

RN [1] ?
 RP SEQUENCE FROM N.A.
 RA Strauberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC027080; AAH27080.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match 25.5%; Score 1739; DB 11; Length 367;
 Best Local Similarity 88.0%; Pred. No. 8.2e-125;
 Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 889 MALESILRRRFTTHQSDVSYGVTVWELMTGAKPYDGIAPAREIPDLLEKGERLPDPPICT 948
 DB 1 MALESILRRRFTTHQSDVSYGVTVWELMTGAKPYDGIAPAREIPDLLEKGERLPDPPICT 60
 QY 949 IDVYIMVKCMKIDSECRPRRELVSFSSMARDPQRFVVIQNEDELGPASPLDSTYRSI 1008
 DB 61 IDVYIMVKCMKIDSECRPRRELVSFSSMARDPQRFVVIQNEDELGPASPLDSTYRSI 120
 QY 1009 LEDDDMGDLVDAEEYLVPOQGFPCPDPAAGGVHHRHSSSTRSGGDLTLGLEPSEE 1068
 DB 121 LEDDDMGDLVDAEEYLVPOQGFPCPDPAAGGVHHRHSSSTRSGGDLTLGLEPSEE 180
 QY 1069 EAPRSPLAPSGAGSDVFDGDLGMAAKGLOSPLTHDPSFLQRYSEDPVPLPSETDGY 1128
 DB 181 EAPRSPLAPSGAGSDVFDGDLGMAAKGLOSPLTHDPSFLQRYSEDPVPLPSETDGY 240
 QY 1129 APLTCSQPEYVNOVDVPPQPSREGPLPAAPAGATLERAKTLSGKXGVYKDYFAFG 1188
 DB 241 APLTCSQPEYVNOVDVPPQPSREGPLPAAPAGATLERAKTLSGKXGVYKDYFAFG 300
 QY 1189 GAVENPEYLTPQGAAPQPPAPFAFDNLYYWDQDPPREGAPSTFGTPTAENPEY 1248
 DB 301 GAVENPEYLTPQGAAPQPPAPFAFDNLYYWDQDPPREGAPSTFGTPTAENPEY 360
 QY 1249 LGLDVPPV 1255
 DB 361 LGLDVPPV 367

RESULT 10

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1] ?
 RP SEQUENCE FROM N.A.
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
 RT "Novel human cDNA clones with function of inhibiting cancer cell
 RT growth."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF318349; AAL55856.1; -
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR InterPro; IPR004019; YLP_motif.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF02757; YLP; 2.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 KW Hypothetical protein.

SQ SEQUENCE 412 AA; 44702 MW; 034397FE32702BC CRC64;
 Query Match 24.9%; Score 1697.5; DB 4; Length 412;
 Best Local Similarity 80.5%; Pred. No. 1.5e-121;
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 889 MALESILRRRFTTHQSDVSYGVTVWELMTGAKPYDGIAPAREIPDLLEKGERLPDPPICT 948
 DB 1 MALESILRRRFTTHQSDVSYGVTVWELMTGAKPYDGIAPAREIPDLLEKGERLPDPPICT 60
 QY 949 IDVYIMVKCMKIDSECRPRRELVSFSSMARDPQRFVVIQNEDELGPASPLDSTYRSI 1008
 DB 61 IDVYIMVKCMKIDSECRPRRELVSFSSMARDPQRFVVIQNEDELGPASPLDSTYRSI 120
 QY 1009 LEDDDMGDLVDAEEYLVPOQGFPCPDPAAGGVHHRHSSSTRSGGDLTLGLEPSEE 1068
 DB 121 LEDDDMGDLVDAEEYLVPOQGFPCPDPAAGGVHHRHSSSTRSGGDLTLGLEPSEE 180
 QY 1069 EAPRSPLAPSGAGSDVFDGDLGMAAKGLOSPLTHDPSFLQRYSEDPVPLPSETDGY 1128
 DB 181 EAPRSPLAPSGAGSDVFDGDLGMAAKGLOSPLTHDPSFLQRYSEDPVPLPSETDGY 240
 QY 1129 APLTCSQPEYVNOVDVPPQPSREGPLPAAPAGATLERAKTLSGKXGVYKDYFAFG 1188
 DB 241 APLTCSQPEYVNOVDVPPQPSREGPLPAAPAGATLERAKTLSGKXGVYKDYFAFG 300
 QY 1189 GAVENPEYLTPQGAAPQPPAPFAFDNLYYWDQDPPREGAPSTFGTPTAENPEY 1248
 DB 301 GAVENPEYLTPQGAAPQPPAPFAFDNLYYWDQDPPREGAPSTFGTPTAENPEY 360
 QY 1221 YVMDQDPPREGAPSTFGTPTAENPEY 1245
 DB 361 YVMDQDPPREGAPSTFGTPTAENPEY 410

RESULT 11

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1] ?
 RP SEQUENCE FROM N.A.
 RA Johnson A., Beug H.;
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities."
 RL Oncogene 9:1307-1320(1994).
 DR EMBL; S69372; AAC60725.1; -
 DR HSSP; P03322; 1A6S.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR004028; Retro_M.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF02813; Retro_M; 1.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR ATP_binding; Tyrosine-kinase; Tyrosine-protein kinase.
 SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EFB1D63 CRC64;

Query Match 23.6%; Score 1609; DB 15; Length 729;
 Best Local Similarity 52.5%; Pred. No. 2.1e-114;

RX MEDLINE=90206603; PubMed=1969616;
RA Brunkin A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the
RT transforming potential of the oncogene v-erb-B".
RL Oncogene 5:15-24(1990).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
-1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; X52209; CA36459.1; -.
DR HSSP; P10828; ZML.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Steroid_receptor.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR001628; Znf_Casteroid.
DR Pfam; PF00104; hormone_rec.1.
DR Pfam; PF00069; kinase.1.
DR Pfam; PF00105; zf-C4.1.
DR PRINTS; PR00398; STRDHOMOMER.
DR PRINTS; PR00447; STROIDFINGER.
DR ProDom; PD000001; Euk_pkinase.1.
DR ProDom; PD000035; Znf_Casteroid.1.
DR SMART; SM00430; HOL1.1.
DR SMART; SM00219; Tyrc; 1.
DR SMART; SM00399; Znf_C4.1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR.1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP.1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM.1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR.1.
KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Transferase; Tyrosine-protein kinase;
KW Zinc-finger.
SQ
SEQUENCE 962 AA: 108320 MW: 3C5AED791BE95CE CRC64;
Query Match 22.7%; Score 1547.5; DB 15; Length 962;
Best Local Similarity 49.7%; Pred. No. 1.6e-109;
Matches 346; Conservative 74; Mismatches 149; Indels 127; Gaps 20;
QY 541 VEEGRVQLGPRE-YVNAR-HCLP-----CHPEQ 568
DB 354 IEKCOESTLAFEHYINRKANHIFMSKLMKVADLMGAYHASFLLMKVCEPTELS 413
QY 569 FONGSVTCFGEADQVACAHYKDPFCVACPSGVKPDLSYPMKFPDEGACQPCPI 628
DB 414 PQE-----VGP--DHCMKCAHFIDGPHCVKACPAVGLGENDTL-VKXYADANAVCQLCHP 465
QY 629 NCHTSQVLDLKGCPAEGRASPLTISVAVV-GILLVVVGVVFGILIKRQKIRKYM 687
DB 466 NCTRGCKGPGLEGCP--NGSKTPSIAAGVVGGLCLVVGIGIGLYLRER-HIVKRTL 521
QY 688 RRLQETELVEPLTPSGAMPQAOIRILKETELRKVKVLSGSAFGFNNFTVSFWLKVPKY 747
DB 522 RRLQERELVEPLTPSGARPQAHRLIKETEPKVKVGLGFGAFG---TVYGLMTIFPG 577
QY 748 SASHELV---LRENTSPKANKELIDEAYVMAGVSPYVSRLLIGICTSTVQLTOLMPY 803
DB 578 EKVITPVAIKELRATSPKANKELIDEAYVMASVDNPHVCRLLIGICTSTVQLITOLMPY 637
QY 804 GCLLDHVENGRGSGODLLNMCQIAKMSYLEDEVLVHRDLAARNVLKSNHVKIND 863
DB 638 GCLLDYIREHKDNGSQYLLNMCQIAKMGNYLEERIMVHRDLAARNVLKTPHOKXITD 697
QY 864 FGLARLLDIDETEVHADGKVPKIMMALESILRRRFTHOSDWSYGVYVWELMTFGAKPY 923
DB 698 FGLAKOLGADDEKEVHAAGKVPKIMMALESILHRIYTHOSDWSYGVYVWELMTFGSKPY 757
QY 924 DGIPAREIPDLKEXGERLPOPICTIVYMIYKCMWIDSECPREPRVSESRNARPP 983
DB 758 DGIPASEISSVLEKGERLPOPICTIVYMIYKCMWISGADSPKRELLAERSKARPP 817
QY 984 QRVVVQ-NEDLGASPLSTFYRSLDEDDMDGLVDAEYLVPOOGFFCPDPAPGAGN 1042
DB 818 PRVLTIGDERMHLPSPTDSKFFYRTLMEEDMEDIVDAEYLVPHQGF----- 866

QY 1043 VHHRRSSRSGGGDTLLGLBSEEEBAPRPLAPSGAGSDVFDGLGGAAGLQSLP 1102
DB 867 -----NSPST-----STPLLSLSASNS-----NSATKCIDRNG 896
QY 1103 THDSPLORISEDTVLBETDGYVAPLTCSQPEPVYVNPDPVPPQPSREGPLAARP 1162
DB 897 GH-----PVREGFL-----PAPEVNG-LMPKKFSTAVYQNOIYNY 932
QY 1163 AGAT-LERAKTSPGKNVYKDVAFGAVENPEYL 1197
DB 933 ISLTATSKLPMDSRYQN-----SHSTAVDNEPYL 961
RESULT 14
Q85468 PRELIMINARY; PRT; 545 AA.
AC Q85468;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Avian Erythroblastosis virus (T834) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=2897102;
RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
RT virus mutants that are temperature sensitive for transformation".
RL Oncogene Res. 1:265-278(1987).
DR EMBL; X06943; CA30024.1; -.
DR HSSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; Pkinase.1.
DR ProDom; PD000001; Euk_pkinase.1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP.1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM.1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR.1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ
SEQUENCE 545 AA: 60899 MW: 140DCE8C6CAF8AF4 CRC64;
Query Match 22.6%; Score 1539; DB 15; Length 545;
Best Local Similarity 52.7%; Pred. No. 3.1e-109;
Matches 333; Conservative 71; Mismatches 128; Indels 100; Gaps 17;
QY 578 GPEADQVACAHYKDPFCVACPSGVKPDLSYPMKFPDEGACQPCPICTHSCVLD 637
DB 1 GP--DHCMKCAHFIDGPHCVKACPAVGLGENDTL-VKXYADANAVCQLCHPCTRGCKGP 57
QY 638 DDKCPAEGRASPLTISVAVV-GILLVVVGVVFGILIKRQKIRKYMRLQETEL 696
DB 58 GLEGCP--NGSKTPSIAAGVVGGLCLVVGIGIGLYLRER-HIVKRTLRLQEREL 113
QY 697 VEPITPSGAMPQAOIRILKETELRKVKVLSGSAFGFNNFTVSFWLKVPKYVKSASHELV-- 754
DB 114 VEPITPSGAPNQHMLILKETEPKVKVGLGFGAFG---TVYGLMTIFEGEKVITPVAI 169
QY 755 -LRENTSPKANKELIDEAYVMAGVSPYVSRLLIGICTSTVQLTOLMPYGLDHVE 812
DB 170 KELREATSPKANKELIDEAYVMASVDNPHVCRLLIGICTSTVQLITOLMPYGLDYIRE 229
QY 813 NRGRLSGODLLNMCQIAKMSYLEDEVLVHRDLAARNVLKSNHVKITDGLARLDI 872
DB 230 HKDNGSQYLLNMCQIAKMGNYLEERIMVHRDLAARNVLKTPHOKXITDGLAQLCA 289
QY 873 DETEVHADGKVPKIMMALESILRRRFTHOSDWSYGVYVWELMTFGAPYDGIAPREIP 932
DB 290 DEKRYHAGKVPKIMMALESILHRIYTHOSDWSYGVYVWELMTGSKPYDGIAPASEIS 349

FT	Region	/label= Cysteine_rich_domain
FT	Region	210..724
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	250..264
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Domain	324..483
FT	Domain	/label= ligand_binding_domain
FT	Region	325..339
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	369..383
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	465..479
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Domain	484..623
FT	Domain	/label= Cysteine_rich_domain
FT	Region	579..593
FT	Region	/label= insertion_region
FT	Domain	/note= "suitable for foreign epitope insertion"
FT	Domain	624..654
FT	Region	/label= Transmembrane_domain
FT	Region	632..652
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	653..667
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Domain	655..1010
FT	Domain	/label= Tyrosine_kinase_domain
FT	Region	661..675
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	695..709
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	710..730
FT	Region	/label= insertion_region
FT	Domain	/note= "suitable for foreign epitope insertion"
FT	Domain	1011..1235
FT	Domain	/label= C-terminal_domain
XX		
PN		WC2000020027-A2.
XX		
PD	13-APR-2000.	
XX		
PF	05-OCT-1999;	99WC-DK00525.
XX		
XX	05-OCT-1998;	98DK-0001261.
PR	20-OCT-1998;	98US-0105011.
XX		
XA	(MEBI-) M & E BIOTECH AS.	
XX		
PI	Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;	
PI	Gautam A, Birk P, Karlsson G;	
XX		
DR	WPI; 2000-349917/30.	
XX		
XX	N-PSDB; AAA09455.	
XX		
PT	Including immune responses to weakly immunogenic, tumor associated	
PT	peptide antigens for the treatment of breast and prostate cancer	
XX		
PS	Claim 62; Page 193-198; 220pp; English.	
XX		
CC	This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of	
CC	Her2 can be used in the claimed method as an autovaccine to induce a CTL	
CC	response. Subdominant CTL epitopes/ antibody binding regions and	
CC	cysteine residues involved in disulfide bonds are preserved in the	
CC	immunogenized forms. Regions suitable for the insertion of foreign T	
CC	helper epitopes were identified (see features table). The method	

is used for inducing immune responses against weakly immunogenic cell-associated peptide antigens (PA) such as those associated with cancers (self-proteins), e.g. human prostate specific membrane antigen (PSM), herceptin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b). The method comprises effecting simultaneous presentation by antigen producing cells (APCs) of the animals immune system of: (1) at least 1 CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1 B-cell group derived from the cell-associated PA; and (2) at least 1 first T helper cell group which is foreign to the animal. Analogues of human PSM, human Her2 and human/murine FGF8b comprising a substantial part of all known and predicted CTL and B-cell epitopes of the respective PA and including at least one foreign T helper epitope are also claimed. The method is used to treat prostate, prostate/breast or breast cancer when the PA is human PSM, FGF8b and Her2, respectively.

Query Match	97.0%	Score 6608	DB 21	Length 1255	
Best Local Similarity	97.1%	Pred. No. 0			
Matches 1223	Conservative 10	Mismatches 17	Indels 10	Gaps 3	
Qy	1	MELAAICRMGILLALLP	PGCAASTQVCSTDMKRLP	SPASEPETHLDMRLHYOGGCVQVQNTL 60	
Db	1	MELAAICRMGILLALLP	PGCAASTQVCSTDMKRLP	SPASEPETHLDMRLHYOGGCVQVQNTL 60	
Qy	61	ELTYLPTNASLSFL	ODIEOVQSYVLI	AHNQVRQVPLRLRIYRGTOFLFEDNALAYLDNG 120	
Db	61	ELTYLPTNASLSFL	ODIEOVQSYVLI	AHNQVRQVPLRLRIYRGTOFLFEDNALAYLDNG 120	
Qy	121	DLPLNNTPTV	TGASPGGISEL	IELRLSTELIKGCVLIQGNPOLCYODITLIMKDIIFHNKNOLA 180	
Db	121	DLPLNNTPTV	TGASPGGISEL	IELRLSTELIKGCVLIQGNPOLCYODITLIMKDIIFHNKNOLA 180	
Qy	181	LTLIDITNRSRACH	PGSPMC	KGSRGCGESSEDQSLRTTVAGGCAACGKGLPTDCHEQC 240	
Db	181	LTLIDITNRSRACH	PGSPMC	KGSRGCGESSEDQSLRTTVAGGCAACGKGLPTDCHEQC 240	
Qy	241	AAAGCTGPKHSD	CLACHFNHSGICE	LHCALVTYNTDTPESMNPBGRYTFGASCTYACP 300	
Db	241	AAAGCTGPKHSD	CLACHFNHSGICE	LHCALVTYNTDTPESMNPBGRYTFGASCTYACP 300	
Qy	301	YNLTSDVDSCTL	YCNLHNOEYTA	DGRCREKSKCAVCGILGMOVYKANSKRTIGTT 360	
Db	301	YNLTSDVDSCTL	YCNLHNOEYTA	DGRCREKSKCAVCGILGMEHLREVRAVTSAN 360	
Qy	361	ELEFPAGCKKI	FGSLAF	LPESPFGDPASNTAPLQEOQVETLEEITGVLYSAMPDLSL 420	
Db	361	IOEPAGCKKI	FGSLAF	LPESPFGDPASNTALQEOQVETLEEITGVLYSAMPDLSL 420	
Qy	421	DLSPFONLQVIR	RIIRIANGAYS	LTIOGIGISWLGRLRLREGSGLAIHHNTHLCVHTV 480	
Db	421	DLSPFONLQVIR	RIIRIANGAYS	LTIOGIGISWLGRLRLRELGSGLAIHHNTHLCVHTV 480	
Qy	481	PMQOLFNP	HPQALLHTANR	PEDECYGEGLACHQICARHGCMGSGPQCVNCSOFLRGQC 540	
Db	481	PMQOLFNP	HPQALLHTANR	PEDECYGEGLACHQICARHGCMGSGPQCVNCSOFLRGQC 540	
Qy	541	VEBCRYLQGI	PREYVNA	RHCLPCHBECQ	PQNGSVTCGPEADQCVACAHKDPFCVAAC 600
Db	541	VEBCRYLQGI	PREYVNA	RHCLPCHBECQ	PQNGSVTCGPEADQCVACAHKDPFCVAAC 600
Qy	601	PSGVKPDLS	TPMPLMKR	PPDEEGACQCP	INCTHSCVDLDDKGCAPAEQAPASPLTSIVSAVVG 660
Db	601	PSGVKPDLS	TPMPLMKR	PPDEEGACQCP	INCTHSCVDLDDKGCAPAEQAPASPLTSIVSAVVG 660
Qy	661	ILLVVVLGVV	FGILIRGQK	IRKTYMRL	LQETELVEPLTPSGAMPNQAOKRLKETEL 720
Db	661	ILLVVVLGVV	FGILIRGQK	IRKTYMRL	LQETELVEPLTPSGAMPNQAOKRLKETEL 720
Qy	721	RKYKVLGSGA	FGFENFTVS	FWLRVP----	KVSAASHLEVLRENTSPRANKETIDEAYVMA 775
Db	721	RKYKVLGSGA	FGFENFTVS	FWLRVP----	KVSAASHLEVLRENTSPRANKETIDEAYVMA 775

QY 776 GVSGPYVSRLLIGLICITSTVQLVTLQMPYGCGLIDHYENRGRGLGSDLLIMCMQIAKGMVY 835
 DB 776 GVSGPYVSRLLIGLICITSTVQLVTLQMPYGCGLIDHYENRGRGLGSDLLIMCMQIAKGMVY 835
 QY 836 LEDVRLVHRDLAARNLVKSPNHVKITDPGLARLLDIDETEHADGKVPFKMMALSSIL 895
 DB 836 LEDVRLVHRDLAARNLVKSPNHVKITDPGLARLLDIDETEHADGKVPFKMMALSSIL 895
 QY 896 RRRFTHQSDVMSVGTVMELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIM 955
 DB 896 RRRFTHQSDVMSVGTVMELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIM 955
 QY 956 VKCMTIDSECRPFRELVSFESRMARDPQRFVIVIONEDLGASPIIDSTFYRSLLEDMDMG 1015
 DB 956 VKCMTIDSECRPFRELVSFESRMARDPQRFVIVIONEDLGASPIIDSTFYRSLLEDMDMG 1015
 QY 1016 DLVDAEEYLVPOGGFPCPDPAAGAGMWHRRSSSTSGGDLTLGLGEPEEERAPSP 1075
 DB 1016 DLVDAEEYLVPOGGFPCPDPAAGAGMWHRRSSSTSGGDLTLGLGEPEEERAPSP 1075
 QY 1076 APSEGAGSDVFDGDLGMAKGLQSLPTHDPSPILQRYSEDPTVPJPSETDGYVAPLTCSP 1135
 DB 1076 APSEGAGSDVFDGDLGMAKGLQSLPTHDPSPILQRYSEDPTVPJPSETDGYVAPLTCSP 1135
 QY 1136 QPEYVNPQPVRRPQSPRREGSPAPARPAATLEAKTISPGKNGVYKQVFAFGAVENPE 1195
 DB 1136 QPEYVNPQPVRRPQSPRREGSPAPARPAATLEAKTISPGKNGVYKQVFAFGAVENPE 1195
 QY 1196 YLTPQGAAPQPHPPAFSPAFDNLYYMDQDPERGAPSTFKGTPTAENPEYGLDVPV 1255
 DB 1196 YLTPQGAAPQPHPPAFSPAFDNLYYMDQDPERGAPSTFKGTPTAENPEYGLDVPV 1255

RESULT 2
 ID AAE12130 standard; Protein; 1255 AA.
 AC AAE12130;
 DT 18-DEC-2001 (first entry)
 DE Human tyrosine kinase-type receptor, HER-2.
 KW Therapeutic compound; major histocompatibility complex; vaccine;
 KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
 KW adoptive immunotherapy; anti-cancer; breast cancer antigen; Apc;
 KW antigen presenting cell; human; tyrosine kinase-type receptor.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Region 774..782
 FT /note="Antigenic epitope"
 PN MO200168677-A2.
 PD 20-SEP-2001.
 PF 16-MAR-2001; 2001MO-US40328.
 PR 16-MAR-2000; 2000US-0527487.
 XX (GEN2) GENZYME CORP.
 XX Nicolette CA;
 XX WPI; 2001-616284/71.
 XX N-PSDB; AAD19731.
 PT Novel synthetic therapeutic compound for inducing immune response and
 PT for use in adoptive immunotherapy, has enhanced binding to major
 PT histocompatibility molecules and enhanced immunoregulatory properties

XX Claim 4: Page 63-67; 69pp; English.
 PS The invention relates to synthetic therapeutic compounds (antigenic
 CC peptides) with enhanced binding to major histocompatibility complex
 CC (MHC) molecules and enhanced immunoregulatory properties relative
 CC to their natural counterparts. Compounds of the invention are useful
 CC for inducing an immune response in a subject and for use in adoptive
 CC immunotherapy. They are useful as components of anti-cancer vaccines
 CC and to expand immune effector cells that are specific for cancers
 CC characterized by expression of the breast cancer antigen, HER-2.
 CC polynucleotides that encode peptides of the invention are useful as
 CC hybridization probes and as primers for the detection of genes of gene
 CC transcripts that are expressed in antigen presenting cells (APCs), to
 CC confirm transduction of polynucleotides into host cells. The present
 CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
 CC of the invention are designed based on the HER-2 antigenic peptide
 CC (774-782).
 SQ Sequence 1255 AA;
 QY Query Match 97.0%; Score 6608; DB 22; Length 1255;
 QY Best Local Similarity 97.1%; Pred. No. 0;
 QY Matches 1223; Conservative 10; Mismatches 17; Indels 10; Gaps 3;
 QY 1 MELAIQRMGLLALLPQAASTQVCTGTMTKLRLPASPETHLMDLRLUYGCGVQGNL 60
 DB 1 MELAIQRMGLLALLPQAASTQVCTGTMTKLRLPASPETHLMDLRLUYGCGVQGNL 60
 QY 61 ELTYLPTNASTLFLDIOVGQYVLIANQVQVLPQRLRYRGQLFEDNYALAVLDNG 120
 DB 61 ELTYLPTNASTLFLDIOVGQYVLIANQVQVLPQRLRYRGQLFEDNYALAVLDNG 120
 QY 121 DPLNNTPTVQASPGGLRELQRLSTELIKGGLVLIQNPOLCYQDTILMKDIFKNNQLA 180
 DB 121 DPLNNTPTVQASPGGLRELQRLSTELIKGGLVLIQNPOLCYQDTILMKDIFKNNQLA 180
 QY 181 LTLIDNRBRACHPGSPMKGRCMGESSEDDQSLTRTVACGACGACGRLPTCCHEQC 240
 DB 181 LTLIDNRBRACHPGSPMKGRCMGESSEDDQSLTRTVACGACGACGRLPTCCHEQC 240
 QY 241 AAGCTGPKHSDCLACILFNHSGICELHCPALVYNTDTPFESMPNDEGRYTTGASCVTACP 300
 DB 241 AAGCTGPKHSDCLACILFNHSGICELHCPALVYNTDTPFESMPNDEGRYTTGASCVTACP 300
 QY 301 YNLTSDVGSCTLVCPCLHNGQVTAEDGTQRCGKSKPCARVCYGLQWQYIRANSKFIGIT 360
 DB 301 YNLTSDVGSCTLVCPCLHNGQVTAEDGTQRCGKSKPCARVCYGLQWQYIRANSKFIGIT 360
 QY 361 ELFFAGCKKI FGSILAFIPSPDPDPASNTAPLPQROLOVFETLEITGYLYISAMPDLP 420
 DB 361 ELFFAGCKKI FGSILAFIPSPDPDPASNTAPLPQROLOVFETLEITGYLYISAMPDLP 420
 QY 421 DLSVFQNLQVIRGLIHNGAYSLLTQGLGISWLGRLSRLGSGALLIHNTHLCFVHTV 480
 DB 421 DLSVFQNLQVIRGLIHNGAYSLLTQGLGISWLGRLSRLGSGALLIHNTHLCFVHTV 480
 QY 481 PMDQLFRNHQALLHTARPREDECEGEGACGQLCARHGCMPGTQCVNSQFLRGQEC 540
 DB 481 PMDQLFRNHQALLHTARPREDECEGEGACGQLCARHGCMPGTQCVNSQFLRGQEC 540
 QY 541 VEECRVLOGLPREYVNAARCLPCHPECOFQNSVYCFPEADQCAACHYADPFVCVARC 600
 DB 541 VEECRVLOGLPREYVNAARCLPCHPECOFQNSVYCFPEADQCAACHYADPFVCVARC 600
 QY 601 PSGVKPDLSTYMPIMKFPDEEGACQCPINCHSCVDLDDKCPAQASPLTISAVAVG 660
 DB 601 PSGVKPDLSTYMPIMKFPDEEGACQCPINCHSCVDLDDKCPAQASPLTISAVAVG 660
 QY 661 ILVVVVLGVVFGILLKRRQKIRKTYMRLLOETLVEPLTPSGAMPQQAOMRIKETEL 720
 DB 661 ILVVVVLGVVFGILLKRRQKIRKTYMRLLOETLVEPLTPSGAMPQQAOMRIKETEL 720

Db 836 LEDVLRHRLAARNVLKSPNHVKTIDFGIARLLDIDETEHADGKGVKIMWALESL 895
 QY RRRFTHOSDVMSTGYTVELMTFGAKPYDGIAPAEIPDLLEKGRLEQPPICTIDYVMIM 955
 Db RRRFTHOSDVMSTGYTVELMTFGAKPYDGIAPAEIPDLLEKGRLEQPPICTIDYVMIM 955
 QY 956 VKCMMIDSECRPRFELVSEFSRMAADPQRFVVIQNEIDLGPASPLDSTFYRSLIEDDMG 1015
 Db VKCMMIDSECRPRFELVSEFSRMAADPQRFVVIQNEIDLGPASPLDSTFYRSLIEDDMG 1015
 QY 1016 DLVDAEFLVPOGGFCPCDPAPGAGGMVHHRRSSSTRSGGDLTLLEPSEBAPRSP 1075
 Db DLVDAEFLVPOGGFCPCDPAPGAGGMVHHRRSSSTRSGGDLTLLEPSEBAPRSP 1075
 QY 1076 APSEGAGSDVEPDGDLGMAAKGLQSLPTHDPSPLQRYSEDPVLEPSETDGYVAPLTCSP 1135
 Db APSEGAGSDVEPDGDLGMAAKGLQSLPTHDPSPLQRYSEDPVLEPSETDGYVAPLTCSP 1135
 QY 1136 QPEYVNOVDVPRQPSPREGLPAARPGALIERAKTILSPKKNVWVDVAFGAVENPE 1195
 Db 1136 QPEYVNOVDVPRQPSPREGLPAARPGALIERAKTILSPKKNVWVDVAFGAVENPE 1195
 QY 1196 YLTPOGGAAPQHPHPPAFSPADNLYWDQDPBERGAPPSTFKGTPTAENPEYLGLDVPV 1255
 Db 1196 YLTPOGGAAPQHPHPPAFSPADNLYWDQDPBERGAPPSTFKGTPTAENPEYLGLDVPV 1255
 RESULT 4
 AAU74545
 ID AAU74545 standard; Protein; 1255 AA.
 AC AAU74545;
 XX 23-APR-2002 (first entry)
 DE Human HER2 (ErbB2) polypeptide.
 XX Human; ErbB; epidermal growth factor receptor; receptor;
 KM anti-ErbB antibody-maytensinoid conjugate; cancer; tumour; breast; ovary;
 KM stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
 KM thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
 KM glial disorder; astrocytal disorder; hypothalamic disorder;
 KM glandular disorder; macrophagal disorder; epithelial disorder;
 KM stromal disorder; blastocoelec disorder; inflammatory disorder;
 KM angioenic disorder; immunological disorder.
 XX Homo sapiens.
 OS Homo sapiens.
 PN US200201587-A1.
 PD 03-JAN-2002.
 PF 16-MAR-2001; 2001US-0811123.
 PR 16-MAR-2000; 2000US-189844P.
 PR 05-OCT-2000; 2000US-238327P.
 PA (ERIC/) ERICKSON S.
 PA (SCHM/) SCHWALL R.
 PA (SLIW/) SLIWOMSKI M.
 PI Erickson S, Schwall R, Slivkowski M,
 DR WPI: 2002-163686/21.
 DR N-PSDB; ABK14058.
 PT Treating tumour characterised by overexpression of epidermal growth
 PT factor receptor, ErbB or cancer in mammal, comprises administering
 PT anti-ErbB antibody-maytensinoid conjugate to the mammal
 PS Example 3; Fig 7; 93pp; English.
 CC The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (ErbB) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytensinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses ErbB2. The method is also useful for treating
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
 CC epithelial, stromal, blastocoelec, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)
 CC polypeptide of the invention.
 XX
 SQ Sequence 1255 AA;
 Query Match 97.0%; Score 6608; DB 23; Length 1255;
 Best Local Similarity 97.1%; Pred. No. 0;
 Matches 1223; Conservative 10; Mismatches 17; Indels 10; Gaps 3;

QY 1 MELAALCRWGLLLALPPGAASVCTGTDMKRLPASPTHLDMLSHLYGCGVVGNTL 60
 Db 1 MELAALCRWGLLLALPPGAASVCTGTDMKRLPASPTHLDMLSHLYGCGVVGNTL 60
 QY 61 ELTYLPTNASLFIQDIQEVGVYLIANOVROYPLQRLATVGTQCFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLFIQDIQEVGVYLIANOVROYPLQRLATVGTQCFEDNYALAVLDNG 120
 QY 121 DPLNNTTPTVGTASPGGLREIQRLSTELIKGVLIQGNPOLCYODTLMDIFHKNNQLA 180
 Db 121 DPLNNTTPTVGTASPGGLREIQRLSTELIKGVLIQGNPOLCYODTLMDIFHKNNQLA 180
 QY 181 LTLIDTNRSAQHPGSPKCKGSRWGBSSSEDCQSLRTTVACAGCARKGRLPTDCCHEQC 240
 Db 181 LTLIDTNRSAQHPGSPKCKGSRWGBSSSEDCQSLRTTVACAGCARKGRLPTDCCHEQC 240
 QY 241 AAGCTGPRHSDCLACIHNHSGICELHCPALVTNTDTPFSMPBERRYTFGASCVTACP 300
 Db 241 AAGCTGPRHSDCLACIHNHSGICELHCPALVTNTDTPFSMPBERRYTFGASCVTACP 300
 QY 301 YNYLSTDVSGCTLVCPHNGEVTAEADGTQRCCKSKPCARVCYGLGMQYIKANSKFIGIT 360
 Db 301 YNYLSTDVSGCTLVCPHNGEVTAEADGTQRCCKSKPCARVCYGLGMQYIKANSKFIGIT 360
 QY 361 ELERFAGCKKIRGSLAFIPESFDGSPASNTAPLOEQOVETTELLEITGYLISAWPDSLP 420
 Db 361 IQEFAGCKKIRGSLAFIPESFDGSPASNTAPLOEQOVETTELLEITGYLISAWPDSLP 420
 QY 421 DLSPQNTQVIRGRILNHGAYSLTLQIGISWLGRLRLREISGLALIHNTTHLCFVHTV 480
 Db 421 DLSPQNTQVIRGRILNHGAYSLTLQIGISWLGRLRLREISGLALIHNTTHLCFVHTV 480
 QY 481 PWDQLFRNPHQALLHTANRPEDECVGBGLACHQICARGHCWGPPTCCVNCOSQTLRQEC 540
 Db 481 PWDQLFRNPHQALLHTANRPEDECVGBGLACHQICARGHCWGPPTCCVNCOSQTLRQEC 540
 QY 541 VEECRVLOGLPREVYNARHCPCPECOQONSGVTCGPPEADQCVAAHAKYDPFCVARG 600
 Db 541 VEECRVLOGLPREVYNARHCPCPECOQONSGVTCGPPEADQCVAAHAKYDPFCVARG 600
 QY 601 PSGVPRDLSYMPIMKFPDEGACQPCPINCTHSCVDLDDKCPAQEAPASPLTISAVNG 660
 Db 601 PSGVPRDLSYMPIMKFPDEGACQPCPINCTHSCVDLDDKCPAQEAPASPLTISAVNG 660
 QY 661 ILLVVLGVVFGLILKROQKIRYTRBRILQETELVEPLTPSGAMNOMRLKETEL 720
 Db 661 ILLVVLGVVFGLILKROQKIRYTRBRILQETELVEPLTPSGAMNOMRLKETEL 720
 QY 721 RKVKVLGSGAGCFNNFTVSWLRVP-----KVSASHLEVLRENTSPPRANKELIDEAYVMA 775
 Db 721 RKVKVLGSGAGCFNNFTVSWLRVP-----KVSASHLEVLRENTSPPRANKELIDEAYVMA 775
 QY 776 GVGSPEYRRLGICLTISIVQLVTLQMLFRYGLLDHVRNBRRLSGDLNMCQOIAKMSY 835

```

Db      776 GVGSPIVSRLLGLCTSTVQLTQMLPVGCLLDHVRNRGLSGQDILLNWMQIAKMSY 835
Qy      836 LEDVRLVHRDLAARVLYKSPNHVITDFFGLARLLDIDETEHADGGKVPKXWALESTL 895
Db      836 LEDVRLVHRDLAARVLYKSPNHVITDFFGLARLLDIDETEHADGGKVPKXWALESTL 895
Qy      896 RRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVTMIM 955
Db      896 RRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVTMIM 955
Qy      956 VKCMTIDSECRPRFELVSEFSRMAPDQRFVYVIONEDLGPASPLDSTFFYSLLDEDDMG 1015
Db      956 VKCMTIDSECRPRFELVSEFSRMAPDQRFVYVIONEDLGPASPLDSTFFYSLLDEDDMG 1015
Qy      1016 DLVDAEEYLVPOQGFPCDPAPAGAGMYHHRRSSSTSSGGQDITLGLPESEEPAPRSPL 1075
Db      1016 DLVDAEEYLVPOQGFPCDPAPAGAGMYHHRRSSSTSSGGQDITLGLPESEEPAPRSPL 1075
Qy      1076 ABSEGAGSDVFDGDDGMAKXGLQSLPHDPSPLQRYSEDDTVPLPSTDGVAPLTCSP 1135
Db      1076 ABSEGAGSDVFDGDDGMAKXGLQSLPHDPSPLQRYSEDDTVPLPSTDGVAPLTCSP 1135
Qy      1136 QPEYVNPDPVRPQPSPREGRLPAARPAATLERAKTILSPGNQGVXDVFAGAVENPE 1195
Db      1136 QPEYVNPDPVRPQPSPREGRLPAARPAATLERAKTILSPGNQGVXDVFAGAVENPE 1195
Qy      1196 YLTPQGGAAPQHPHPAPAFADNLTYMDQDPPEGAPSTFKGPTAENPEYIGLDVPIV 1255
Db      1196 YLTPQGGAAPQHPHPAPAFADNLTYMDQDPPEGAPSTFKGPTAENPEYIGLDVPIV 1255

RESULT 5
AA001111
ID AA001111 standard; protein; 1255 AA.
XX
AC AA001111;
DT 01-JAN-1997 (first entry)
XX
DE HER-2/neu protein.
XX
KM HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KM breast cancer; ovary cancer; colon cancer; lung cancer;
KM prostate cancer; immunisation; tumour; vaccine; vector.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 676..1255
FT /label= Intracellular_domain
FT /note= "claimed domain, useful for immunisation"
XX
FN WO9630514-A1.
PD 03-OCT-1996.
XX
PF 28-MAR-1996; 96WO-US01689.
PR 31-MAR-1995; 95US-0414417.
XX
PA (UNIW ) UNIV WASHINGTON.
PI Cheever MA, Distis ML;
XX
DR WPI; 1996-455361/45.
DR N-PSDB; AAT40739.
XX
PT DNA encoding HER-2/neu poly:peptide(s) - used for prevention or
PT treatment of malignancies with which the HER-2/neu oncogene is
PT associated
XX
PS Claim 2; Page 56-61; 71pp; English.

```

```

XX      CC Human HER-2/neu protein (AA001111), also called p185 or c-erbB2, is
CC      CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC      CC protein is over-expressed in various cancers, including breast,
CC      CC the ovarian, colon, lung and prostate. The intracellular domain of the
CC      CC protein can be used to immunise an animal against a malignancy with
CC      CC which the oncogene is associated. The polypeptide can be produced
CC      CC in transformed host cells for use in immunisation. Alternatively,
CC      CC animal cells are transfected in vivo or ex vivo with a viral vector
CC      CC that directs expression of the polypeptide.
XX
SQ      Sequence 1255 AA;
Query Match 96.9%; Score 6602; DB 17; Length 1255;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 1221; Conservative 11; Mismatches 16; Indels 10; Gaps 3;
Qy      1 METALACWGLLALLPFGAASVQVCTGTDMKRLPASPTHLDMLRHLYOGCVVQGNL 60
Db      1 METALACWGLLALLPFGAASVQVCTGTDMKRLPASPTHLDMLRHLYOGCVVQGNL 60
Qy      61 ELTYIPTNASLFLQDIOENQGVYLIANOVROVPLRLRIVRGTOIFEDNYALAVDNG 120
Db      61 ELTYIPTNASLFLQDIOENQGVYLIANOVROVPLRLRIVRGTOIFEDNYALAVDNG 120
Qy      121 DPLNNTPTVGTASPGGLREIQRLSLTEILKGVLIQENPOLCYQDITLMDIFHKNNQLA 180
Db      121 DPLNNTPTVGTASPGGLREIQRLSLTEILKGVLIQENPOLCYQDITLMDIFHKNNQLA 180
Qy      181 LTLIDTNRBRACHPCSPCKSGSRCKWESSDDCGSLTRTVACGACARCKPLPTCCHEQC 240
Db      181 LTLIDTNRBRACHPCSPCKSGSRCKWESSDDCGSLTRTVACGACARCKPLPTCCHEQC 240
Qy      241 AAGCTGPKRSDCLACLHNHSGICELHCPALVTNTDFFESMPNPEGRTFGASCVTACP 300
Db      241 AAGCTGPKRSDCLACLHNHSGICELHCPALVTNTDFFESMPNPEGRTFGASCVTACP 300
Qy      301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGNOYIKANSKFTGIT 360
Db      301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGNOYIKANSKFTGIT 360
Qy      361 ELDFAGCKKIFGSLAFPESEFDGDSASTATLOPQLOVEFTLEITGYLYISAMPDLP 420
Db      361 ELDFAGCKKIFGSLAFPESEFDGDSASTATLOPQLOVEFTLEITGYLYISAMPDLP 420
Qy      421 DLSVFQNLQVIRGRLIHNGAYSILTQGLGISWLGRLSRLRELSGLALIHNTTHLCFVHTV 480
Db      421 DLSVFQNLQVIRGRLIHNGAYSILTQGLGISWLGRLSRLRELSGLALIHNTTHLCFVHTV 480
Qy      481 PMDQLFRNHQALLHTANRPEDECVGEGLACHQUCARHCWGPGPTQCVNCSOFLRQEC 540
Db      481 PMDQLFRNHQALLHTANRPEDECVGEGLACHQUCARHCWGPGPTQCVNCSOFLRQEC 540
Qy      541 VEECRVLOGLPREYVNAARCLPECHPECPQWNSVTCFPEADQCVACAHYDPPFCVARC 600
Db      541 VEECRVLOGLPREYVNAARCLPECHPECPQWNSVTCFPEADQCVACAHYDPPFCVARC 600
Qy      601 PSGVPRDLSYMPKWPDEBGAQCPQPINCTHSCVDLDDKCGAPQASPLTISVAVNG 660
Db      601 PSGVPRDLSYMPKWPDEBGAQCPQPINCTHSCVDLDDKCGAPQASPLTISVAVNG 660
Qy      661 ILLVVLGVVPGILIKRQCKIRKTYMRLLOETLVEPLTPSGAMPQAOQRIILKETEL 720
Db      661 ILLVVLGVVPGILIKRQCKIRKTYMRLLOETLVEPLTPSGAMPQAOQRIILKETEL 720
Qy      721 RKVKVLGSAFGFNNFTVSFWLRVP-----KVSASLLEVLRENTSPKANKELDEAYYMA 775
Db      721 RKVKVLGSAFGF-----TYKGIWIPDEGEVKKIPVA-IKVLRENTSPKANKELDEAYYMA 775
Qy      776 GVGSPIVSRLLGLCTSTVQLTQMLPVGCLLDHVRNRGLSGQDILLNWMQIAKMSY 835
Db      776 GVGSPIVSRLLGLCTSTVQLTQMLPVGCLLDHVRNRGLSGQDILLNWMQIAKMSY 835

```

QY 836 LEDVLRVLRDLAARNVLVKSPNHVKTIDFGIARLLDIDETEHADGGKVPDKMMALESIL 895
 DB 836 LEDVLRVLRDLAARNVLVKSPNHVKTIDFGIARLLDIDETEHADGGKVPDKMMALESIL 895
 QY 896 RRRFTHQSDVMSYGVTTWELMTFGAKPYDGI:PAEIPDLLEKSGRLNQPPICITIDVYIM 955
 DB 896 RRRFTHQSDVMSYGVTTWELMTFGAKPYDGI:PAEIPDLLEKSGRLNQPPICITIDVYIM 955
 QY 956 VACWMIIDSECRPRFELVSEFSRMAKDQRFVVIQNEDELGASPLDSTFYRSLLEDDMG 1015
 DB 956 VACWMIIDSECRPRFELVSEFSRMAKDQRFVVIQNEDELGASPLDSTFYRSLLEDDMG 1015
 QY 1016 DLVDAEYLVPQGFCCDPAPAGAGWVHRHRSSTRSGGDI:TLGLEPSEEBAPRSPL 1075
 DB 1016 DLVDAEYLVPQGFCCDPAPAGAGWVHRHRSSTRSGGDI:TLGLEPSEEBAPRSPL 1075
 QY 1076 AASEGASGVFPDGLGMAAKGLOSLPTHDSPLQRYSEDPVLPBETDGYVAPLTCSP 1135
 DB 1076 AASEGASGVFPDGLGMAAKGLOSLPTHDSPLQRYSEDPVLPBETDGYVAPLTCSP 1135
 QY 1136 QPEYVNPDPVRPQPSPREGSLPAARPAATLERAKTLPQKNGVVDVPAFGGAVENPE 1195
 DB 1136 QPEYVNPDPVRPQPSPREGSLPAARPAATLERAKTLPQKNGVVDVPAFGGAVENPE 1195
 QY 1196 YLTPQGAAPQHPAPPASPAFDNIYYNDQPPERGAPESTFKGTPTAENPEYIGLDVPY 1255
 DB 1196 YLTPQGAAPQHPAPPASPAFDNIYYNDQPPERGAPESTFKGTPTAENPEYIGLDVPY 1255
 RESULT 6
 AAM92406
 ID AAM92406 standard; Protein; 1255 AA.
 AC AAM92406;
 XX
 DT 21-APR-1999 (first entry)
 XX
 DE Human HER-2/neu oncogene protein.
 XX
 KM HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
 XX malignancy; treatment; tumour.
 OS Homo sapiens.
 PH Key Location/Qualifiers
 FT Region 676..1255
 FT /note= "region which elicits immune response"
 XX
 PN US869445-A.
 PD 09-FEB-1999.
 XX
 PF 01-APR-1996; 96US-0625101.
 XX
 PR 01-APR-1996; 96US-0625101.
 PR 17-MAR-1993; 93US-003644.
 PR 12-AUG-1993; 93US-0106112.
 PR 31-MAR-1995; 95US-0414417.
 XX
 PA (UNITV) UNITV WASHINGTON.
 PI Cheever MA, Disis ML;
 XX
 DR MPI: 1999-152835/13.
 DR N-PSDB; AAX01912.
 XX
 PT Use of HER-2/neu polypeptides - for eliciting an immune response to
 PT an HER-2/neu associated malignancy, particularly for treating or
 PT preventing tumours
 XX
 PS Claim 3: Column 31-38; 26pp; English.
 XX
 CC This sequence represents the human HER-2/neu oncogene protein. A fragment

CC of this protein is used in a method for eliciting or enhancing an immune
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
 CC B cells to produce an immune response to the HER-2/neu protein. The
 CC method can be used for immunisation against a malignancy in which the
 CC HER-2/neu oncogene is associated and in the treatment of an existing
 CC tumour, or to prevent tumour occurrence or recurrence.
 CC
 XX

Sequence 1255 AA;

Query Match 96.9%; Score 6602; DB 20; Length 1255;

Best Local Similarity 96.9%; Pred. No. 0;

Matches 1221; Conservative 11; Mismatches 18; Indels 10; Gaps 3;

QY 1 MELALCRMGLLALPPGAASVCTGTDMKRLPASPETHLMDLRLHYQCCVVOGNTL 60
 DB 1 MELALCRMGLLALPPGAASVCTGTDMKRLPASPETHLMDLRLHYQCCVVOGNTL 60
 QY 61 ELTYLPTNASLSPLODIOEVGYVLIANOVROVPLRLRVRGTQLFEDNYALAVLDNG 120
 DB 61 ELTYLPTNASLSPLODIOEVGYVLIANOVROVPLRLRVRGTQLFEDNYALAVLDNG 120
 QY 121 DELNNTTPTGASPGGLREQLRLSLTEIKGGLVLIQNPOLCYODTILMKDIFHXNNQLA 180
 DB 121 DELNNTTPTGASPGGLREQLRLSLTEIKGGLVLIQNPOLCYODTILMKDIFHXNNQLA 180
 QY 181 LTLIDTNRSRACHPCSPKCKSGSRCKWSESSDDCSLTTRTVCAAGCARGKGLPTDCCHEQC 240
 DB 181 LTLIDTNRSRACHPCSPKCKSGSRCKWSESSDDCSLTTRTVCAAGCARGKGLPTDCCHEQC 240
 QY 241 AAGCTGPHNSDCLACLFHNHSGICELCPALVTYNTDTFESMPREERYTFGASCTVACP 300
 DB 241 AAGCTGPHNSDCLACLFHNHSGICELCPALVTYNTDTFESMPREERYTFGASCTVACP 300
 QY 301 YNYLSTDVSGCTLVCPILHNOEVTAEADGTORCEKSKPCARVYGLGNOYIKANSKFTGIT 360
 DB 301 YNYLSTDVSGCTLVCPILHNOEVTAEADGTORCEKSKPCARVYGLGNOYIKANSKFTGIT 360
 QY 361 ELLEFAGCKKIFGSLAFLEPSFDGPASTATLOPEQLOVFTTEILGYLYISMPDLSLP 420
 DB 361 ELLEFAGCKKIFGSLAFLEPSFDGPASTATLOPEQLOVFTTEILGYLYISMPDLSLP 420
 QY 421 DLSVFQNLQVIRGRILHNGAVSLTLQGLGISLWLGRLSRLSGSLALIHNTHLCFVHTV 480
 DB 421 DLSVFQNLQVIRGRILHNGAVSLTLQGLGISLWLGRLSRLSGSLALIHNTHLCFVHTV 480
 QY 481 PWDQLFKNPQALHTTANRPDECEVGBGLACHQICARHCWGPPTQCVNCSOFLRQEC 540
 DB 481 PWDQLFKNPQALHTTANRPDECEVGBGLACHQICARHCWGPPTQCVNCSOFLRQEC 540
 QY 541 VEEGRVLOGLPREYVNAHCLPCHPECCOPONGSVTCGPEADQCVACAHYNDPFCVARC 600
 DB 541 VEEGRVLOGLPREYVNAHCLPCHPECCOPONGSVTCGPEADQCVACAHYNDPFCVARC 600
 QY 601 PSGVXPDLSTYPMKFPDEBAGACPCPIINCTHSCVDLDDKGCPEAGBASPLTISAVVG 660
 DB 601 PSGVXPDLSTYPMKFPDEBAGACPCPIINCTHSCVDLDDKGCPEAGBASPLTISAVVG 660
 QY 661 ILVVVTLGVVREGILILKROQKIRKXTVRLLOEELVPELPLPSGAMNOMQRLTXETEL 720
 DB 661 ILVVVTLGVVREGILILKROQKIRKXTVRLLOEELVPELPLPSGAMNOMQRLTXETEL 720
 QY 721 RKVKYLGSGAFGFNNFTVSFWLRVP----KVSASHLEVIRENTSPRANKEILDEAYVMA 775
 DB 721 RKVKYLGSGAFGF---TYKGIWIPDGENVXIPVA-IKVLRENTSPRANKEILDEAYVMA 775
 QY 776 GVGSPYVSRLLGICLTSTVQLVTQGLMPGCLLDHYRENRGLSGQDILNMCQIAKMSY 835
 DB 776 GVGSPYVSRLLGICLTSTVQLVTQGLMPGCLLDHYRENRGLSGQDILNMCQIAKMSY 835
 QY 836 LEDVLRVLRDLAARNVLVKSPNHVKTIDFGIARLLDIDETEHADGGKVPDKMMALESIL 895
 DB 836 LEDVLRVLRDLAARNVLVKSPNHVKTIDFGIARLLDIDETEHADGGKVPDKMMALESIL 895

DB 956 VKCWMIDSECRPRRELIVSEFSRMAADPQRFVVIQNEDELGPASPLDSTFYRSLLEDDMG 1015
QY 1016 DLVDAEEYLVPOQGFCCDPAPAGAGVHHRHSSSTRSGGDLTLGLBSEBEARSP 1075
DB 1016 DLVDAEEYLVPOQGFCCDPAPAGAGVHHRHSSSTRSGGDLTLGLBSEBEARSP 1075
QY 1076 APSEGAGSDVDFDGLGMAAKGLQLPTHDSPLOQYSEDPVPLPSETDGYVAPLTCSP 1135
DB 1076 APSEGAGSDVDFDGLGMAAKGLQLPTHDSPLOQYSEDPVPLPSETDGYVAPLTCSP 1135
QY 1136 OPEYVNOQDVPRPPSPREGPLPAPAPAGATLEAKTSLSPGKGVVWDVAFGAVENPE 1195
DB 1136 OPEYVNOQDVPRPPSPREGPLPAPAPAGATLEAKTSLSPGKGVVWDVAFGAVENPE 1195
QY 1196 YLTPGGGAAPQHPHPPAPAFNDL.YYWDQDPPERGA.PPSTFKGTPTAENPEYLGIDVPY 1255
DB 1196 YLTPGGGAAPQHPHPPAPAFNDL.YYWDQDPPERGA.PPSTFKGTPTAENPEYLGIDVPY 1255

RESULT 8
AAy84780
ID AAy84780 standard; Protein; 1255 AA.
XX
AC AAy84780;
XX
DT 08-AUG-2000 (first entry)
XX
DE Amino acid sequence of the SPICE erbB-2 receptor protein.
XX
KM SPICE erbB-2 receptor protein; cell transformation disorder; cancer;
KM tumor cell proliferation; tissue degeneration; arthropathy;
KM bone resorption; inflammatory disease; degenerative disorder;
KM wound healing.
XX
OS Homo sapiens.
XX
PN WO2000020579-A1.
XX
PD 13-APR-2000.
XX
PF 01-OCT-1999; 99WO-CA00912.
XX
PR 02-OCT-1998; 98US-0165192.
XX
PA (UWMC-) UNIV MCMASTER.
PI Muller WJ, Siegel PM;
XX
DR WPI; 2000-303768/26.
DR N-PSDB; AAA14812.
XX
PT Nucleic acid encoding an erbB 2 receptor protein designated SPICE
PT erbB-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS Claim 3; Fig 2; 60pp; English.
XX
CC The present sequence represents a SPICE erbB-2 receptor protein. The
CC protein has an in-frame deletion of 16 amino acids, 2 of which are
CC conserved cysteine residues, compared to the unspliced protein. The
CC erbB-2 polynucleotide is used to construct probes for detecting
CC disorders of cell transformation such as cancer. Antibodies to the
CC protein may be used to detect SPICE erbB-2 in a sample. Agents
CC (e.g. antisense oligonucleotides) which inhibit the expression of
CC SPICE erbB-2 are useful for reducing tumor cell proliferation and
CC treating cancer. Substances which stimulate SPICE erbB-2 are useful
CC for treating conditions involving damaged cells including conditions
CC in which degeneration of tissue occurs, such as atrophy, bone
CC resorption, inflammatory diseases, degenerative disorders of the
CC central nervous system and wound healing.
XX
SQ Sequence 1255 AA;

Query Match 96.9%; Score 6602; DB 21; Length 1255;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 1221; Conservative 11; Mismatches 18; Indels 10; Gaps 3;
QY 1 MELAALCWMGILLALLPPGASTVCCTGMKRLPASPETHLDMRLHYOGGVQGNL 60
DB 1 MELAALCWMGILLALLPPGASTVCCTGMKRLPASPETHLDMRLHYOGGVQGNL 60
QY 61 ELYLPTNASLFLQDIQEVQGVYLIANOVQVFLQRLRIVRGQLFEDNYALAVLNG 120
DB 61 ELYLPTNASLFLQDIQEVQGVYLIANOVQVFLQRLRIVRGQLFEDNYALAVLNG 120
QY 121 DPLNNTTPTVTSASGGLRELOBSLTETLKGAVLIQNPQLCYQDTLLKMDIFHKNNQLA 180
DB 121 DPLNNTTPTVTSASGGLRELOBSLTETLKGAVLIQNPQLCYQDTLLKMDIFHKNNQLA 180
QY 181 LTLIDTNRSRACHSPKCSKSR.CWGESSEDCOSL.TRTVAGGACARCKPLPTDCHEQC 240
DB 181 LTLIDTNRSRACHSPKCSKSR.CWGESSEDCOSL.TRTVAGGACARCKPLPTDCHEQC 240
QY 241 AAGCTGPHSDCLACLHPNHSICELHCPALVTYNTDFESMPNPEGRTYFGASCYTAAC 300
DB 241 AAGCTGPHSDCLACLHPNHSICELHCPALVTYNTDFESMPNPEGRTYFGASCYTAAC 300
QY 301 YNYLSTDVSGCTLVCPILHNOEVTAEQTCRCKSKPCARVCTGLGKHEHLREVAATYSAN 360
DB 301 YNYLSTDVSGCTLVCPILHNOEVTAEQTCRCKSKPCARVCTGLGKHEHLREVAATYSAN 360
QY 361 ELEFAGCKKIGSLAFIPESFDGPASNTAPLOEQLOVEETLEITGYLYISAMPDLP 420
DB 361 IQEFAGCKKIGSLAFIPESFDGPASNTAPLOEQLOVEETLEITGYLYISAMPDLP 420
QY 421 DLSVFQNTQVIRGRILHNGAYSLTLQGLIGSLGRLBELGSLALIHNTHLCFHTV 480
DB 421 DLSVFQNTQVIRGRILHNGAYSLTLQGLIGSLGRLBELGSLALIHNTHLCFHTV 480
QY 481 PMDOLFRPHQALHTANRPEDECVGSLACHQACARHCMPGPTCCVNCOSFLRQEC 540
DB 481 PMDOLFRPHQALHTANRPEDECVGSLACHQACARHCMPGPTCCVNCOSFLRQEC 540
QY 541 VEECRVLQGLPREVYNARHCLPCHPECOPONGSVTCGPADQCACAHKXDPFCVARC 600
DB 541 VEECRVLQGLPREVYNARHCLPCHPECOPONGSVTCGPADQCACAHKXDPFCVARC 600
QY 601 PSQVRPDLSTYMPIMKFPDEBGACPCPINCCHSCVDLDDXGCPBQASPLTISAVVG 660
DB 601 PSQVRPDLSTYMPIMKFPDEBGACPCPINCCHSCVDLDDXGCPBQASPLTISAVVG 660
QY 661 ILLVVLGVVFGILIKRQCKIRKYTRRLLOETELVEPLTPSGAMNQAOMRILKETEL 720
DB 661 ILLVVLGVVFGILIKRQCKIRKYTRRLLOETELVEPLTPSGAMNQAOMRILKETEL 720
QY 721 RKVKYLSGAGFNNFTVFWLRVP-----KVSASHLEVRENTSPKANEILDEAVMA 775
DB 721 RKVKYLSGAGFNNFTVFWLRVP-----KVSASHLEVRENTSPKANEILDEAVMA 775
QY 776 GVGSPTVSRLLGICLTSTVOLVTLMPYGLLDEVRNRRGLSGODLLNMCQIAKMSY 835
DB 776 GVGSPTVSRLLGICLTSTVOLVTLMPYGLLDEVRNRRGLSGODLLNMCQIAKMSY 835
QY 836 LEDVRLVHRDLAANVIVKSPNHVKITDFGLARLLIDETBYHADGKVPDKMALESIL 895
DB 836 LEDVRLVHRDLAANVIVKSPNHVKITDFGLARLLIDETBYHADGKVPDKMALESIL 895
QY 896 RRRFTHQSDWSYGTATWELMTFAKAYDGPAREIDLEKERRLPQPICTIDVYIM 955
DB 896 RRRFTHQSDWSYGTATWELMTFAKAYDGPAREIDLEKERRLPQPICTIDVYIM 955
QY 956 VKCWMIDSECRPRRELIVSEFSRMAADPQRFVVIQNEDELGPASPLDSTFYRSLLEDDMG 1015
DB 956 VKCWMIDSECRPRRELIVSEFSRMAADPQRFVVIQNEDELGPASPLDSTFYRSLLEDDMG 1015
QY 1016 DLVDAEEYLVPOQGFCCDPAPAGAGVHHRHSSSTRSGGDLTLGLBSEBEARSP 1075

DB 1136 QPEYVNCQDVRRPQPPSPRECGPLPAPAPAGATLERPKTLSPGKNGVVDVFAFGAVENPE 1195
QY 1196 YLTPOGGAAPOPHPAPSPAPFNDLYYDDDDPPRGPAPSPSTFKGTPTAENPEYLGLDVPY 1255
DB 1196 YLTPOGGAAPOPHPAPSPAPFNDLYYDDDDPPRGPAPSPSTFKGTPTAENPEYLGLDVPY 1255

RESULT 10
AAG88267
ID AAG88267 standard: Protein; 1255 AA.

AC AAG88267;
DT 11-SEP-2001 (first entry)
XX
XX
DE HER2/neu amino acid sequence.
KM Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW Immune response; vaccine; tumour; cancer; cytotoxic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
OS Homo sapiens.
XX
XX PN MO200141787-A1.
PD 14-JUN-2001.
XX
XX PF 11-DEC-2000; 2000MO-US33591.
XX
XX PR 10-DEC-1999; 99US-0458299.
XX
XX PA (EPTM-) EPIMONE INC.
PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;
DR WPI: 2001-374995/39.
XX
XX PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer -
PS Disclosure; Page 15; 1999p; English.
XX

CC The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (1)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide sequence
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
CC and immunostimulant activities, and can be used in vaccines. (I), (II)
CC and (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample from a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention.
XX
SQ Sequence 1255 AA;

Query Match 96.9%; Score 6602; DB 22; Length 1255;

Best Local Similarity 96.9%; Pred. No. 0;
Matches 1221; Conservative 11; Mismatches 18; Indels 10; Gaps 3;

QY 1 MELALCWGILLILALPGASTQVCTGTMTKRLPASPETHLDMLRHLYOGCGVQVGNL 60
DB 1 MELALCWGILLILALPGASTQVCTGTMTKRLPASPETHLDMLRHLYOGCGVQVGNL 60
QY 61 ELTYLPTNASLFLQDIOEVQGYLLAHNOYRQVPLQRLRIVRGTOLEFQNALAVLDNG 120
DB 61 ELTYLPTNASLFLQDIOEVQGYLLAHNOYRQVPLQRLRIVRGTOLEFQNALAVLDNG 120
QY 121 DPLNNTTPTVTGASRQGLRELOLRLTELKQGVLCQNPOLCTYDITLMDIFPKNNOLA 180
DB 121 DPLNNTTPTVTGASRQGLRELOLRLTELKQGVLCQNPOLCTYDITLMDIFPKNNOLA 180
QY 181 LTLIDTNSRACHPSPCKSGSRGMSSEDCQSLTRTVACGACGKPLPTDCHEQC 240
DB 181 LTLIDTNSRACHPSPCKSGSRGMSSEDCQSLTRTVACGACGKPLPTDCHEQC 240
QY 241 AAGCTGPRHSDCLALPHNHSIGICELCPALVTYNTDFESMPNREGRYTFGASCVTACP 300
DB 241 AAGCTGPRHSDCLALPHNHSIGICELCPALVTYNTDFESMPNREGRYTFGASCVTACP 300
QY 301 YNYLSTDVSGCTIVGPIHNCQVETADGTQCEKSKPCARVCGLGHEHREVAVTSAN 360
DB 301 YNYLSTDVSGCTIVGPIHNCQVETADGTQCEKSKPCARVCGLGHEHREVAVTSAN 360
QY 361 ELEFAGCKKIFGSLAFPESEFGDPASNTAPLOEQLQVETLEITGYLYISAMPDLSL 420
DB 361 IDEFAGCKKIFGSLAFPESEFGDPASNTAPLOEQLQVETLEITGYLYISAMPDLSL 420
QY 421 DLSVQNIQVIRGRILNHGAYSLTQIGISWGLRLRLRELGSSALIHNTHLCPHNTV 480
DB 421 DLSVQNIQVIRGRILNHGAYSLTQIGISWGLRLRLRELGSSALIHNTHLCPHNTV 480
QY 481 PWDQCFRPHQALHTANRPEDECVGEGLACHQCARHCGMPPTQVNCSPQLRQCEC 540
DB 481 PWDQCFRPHQALHTANRPEDECVGEGLACHQCARHCGMPPTQVNCSPQLRQCEC 540
QY 541 VEECVLQGLPREYVNAHCLPHEPCQPNQSVTCFGEADQCVACAHYKDPFCVARC 600
DB 541 VEECVLQGLPREYVNAHCLPHEPCQPNQSVTCFGEADQCVACAHYKDPFCVARC 600
QY 601 PSGVPEDSLWPMKFPDEBAGACPCPNCHSVDLDKGCPEQASPLTSLVAVNG 660
DB 601 PSGVPEDSLWPMKFPDEBAGACPCPNCHSVDLDKGCPEQASPLTSLVAVNG 660
QY 661 ILVVVLGVVFGIILKRQKIRRYTWRRLLQETELVEPLTPSGAMENQOMRLKETEL 720
DB 661 ILVVVLGVVFGIILKRQKIRRYTWRRLLQETELVEPLTPSGAMENQOMRLKETEL 720
QY 721 RKVKVLGSGAGFNFTVLSFWLRVP-----KVASHLVEURENTSPANKELIDEAVYMA 775
DB 721 RKVKVLGSGAGFNFTVLSFWLRVP-----KVASHLVEURENTSPANKELIDEAVYMA 775
QY 776 GVGSPYVRLGICITSTVQVLTQMPGCLLDVHRENRGLSGQDLNMCQJAKMSY 835
DB 776 GVGSPYVRLGICITSTVQVLTQMPGCLLDVHRENRGLSGQDLNMCQJAKMSY 835
QY 836 LEDVRLVHRDLAARNVLYKSNHYKTDQGLARLLDIDETRYHAGGKVPDKWMALESIL 895
DB 836 LEDVRLVHRDLAARNVLYKSNHYKTDQGLARLLDIDETRYHAGGKVPDKWMALESIL 895
QY 896 RRRPFGHSDVWSYGVTWELMTPAKPYVDGIPAEIIDLKGRILPOPICTIDVYIM 955
DB 896 RRRPFGHSDVWSYGVTWELMTPAKPYVDGIPAEIIDLKGRILPOPICTIDVYIM 955
QY 956 VKCWMIDSECRPRRELIVSEFSRMAPDQRFVVIQNBDLGPASPLDSTFVRSLLDDMG 1015
DB 956 VKCWMIDSECRPRRELIVSEFSRMAPDQRFVVIQNBDLGPASPLDSTFVRSLLDDMG 1015
QY 1016 DLVDAEVLVPOQGFCDPAPAGAGVHRHRSSSTRSGGDTLTLGSESEBAPRSL 1075
DB 1016 DLVDAEVLVPOQGFCDPAPAGAGVHRHRSSSTRSGGDTLTLGSESEBAPRSL 1075

Db 1016 DLVDAEYLVPQGGFFCPDPAAPGAGWVHHRHSSSTRSGGDLTLGLPSEEBEAPRPL 1075
QY 1076 APSEGAGSDVDFDGDIGMAAKGLQSLPTHDPSPICRYSSEDPYVLPBETGGYAAPLTCSP 1135
Db 1076 APSEGAGSDVDFDGDIGMAAKGLQSLPTHDPSPICRYSSEDPYVLPBETGGYAAPLTCSP 1135
QY 1136 QPEYVNOQDVARPQPPSPREGPLPAARFAGATLIERPKTISPQKGVVDVAFGAVENPE 1195
Db 1136 QPEYVNOQDVARPQPPSPREGPLPAARFAGATLIERPKTISPQKGVVDVAFGAVENPE 1195
QY 1196 YLTPQGAAPQHPPPAPSPAFDNLYWDQDPPEBAGPSTFKGTPTAENPEYGLDVPY 1255
Db 1196 YLTPQGAAPQHPPPAPSPAFDNLYWDQDPPEBAGPSTFKGTPTAENPEYGLDVPY 1255
RESULT 11
AAE24067
ID AAE24067 standard; Protein, 1255 AA.
XX
AC AAE24067;
XX
DT 23-SEP-2002 (first entry)
XX
DE Human Her-2 protein.
XX
KW Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KM hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW tumour; gene therapy; phosphorothioate backbone.
XX
OS Homo sapiens.
XX
PN WO200222636-A1.
XX
PD 21-MAR-2002.
XX
PF 12-SEP-2001; 2001WO-US28572.
XX
PR 15-SEP-2000; 2000US-0663834.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Bennett CF, Cowseart LM;
XX
DR WPI; 2002-471192/50.
XX
DR N-PSDB; AAD38904.
XX
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
XX
PT Inflammation or to prevent infection in humans -
XX
PS Example 13; Page 95-107; 116pp; English.
XX
CC The invention relates to antisense compounds targeted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.
XX
SQ Sequence 1255 AA;
Query Match 96.9%; Score 6602; DB 23; Length 1255;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 1221; Conservative 11; Mismatches 18; Indels 10; Gaps 3;

QY 61 ELTYLPTNASLSPLODIOEVQGYVLIANQVQPLQRLIRVGTQLFEDNVALAVLDNG 120
Db 61 ELTYLPTNASLSPLODIOEVQGYVLIANQVQPLQRLIRVGTQLFEDNVALAVLDNG 120
QY 121 DPLNNTPTVTCASFGGLRELOLRSLTEILKGVYLICQNPOLCTQDTILMKDITHKNNQLA 180
Db 121 DPLNNTPTVTCASFGGLRELOLRSLTEILKGVYLICQNPOLCTQDTILMKDITHKNNQLA 180
QY 181 LTLIDTNSRACHPSPCKSGSRCKMGSSESDCOSLTRTYVACGACARCKPLPDDCHEQC 240
Db 181 LTLIDTNSRACHPSPCKSGSRCKMGSSESDCOSLTRTYVACGACARCKPLPDDCHEQC 240
QY 241 AAGCTGPHSDCLACLHNHSGICELHCPALVTYNTDTFESMNPBGRYTFGASCYTACP 300
Db 241 AAGCTGPHSDCLACLHNHSGICELHCPALVTYNTDTFESMNPBGRYTFGASCYTACP 300
QY 301 YNYLSTDVSGCTLVCPLHNOEVTAEDGTQRCCKSKPCARVCTGLGMEHLREYRAVTSAN 360
Db 301 YNYLSTDVSGCTLVCPLHNOEVTAEDGTQRCCKSKPCARVCTGLGMEHLREYRAVTSAN 360
QY 361 ELEFAGCKKIFGSLAFUPSEFSDGPASNTAPLOPEQLOVEFTELEITGYLYISAMPDSLP 420
Db 361 IOEPAGCKKIFGSLAFUPSEFSDGPASNTAPLOPEQLOVEFTELEITGYLYISAMPDSLP 420
QY 421 DLSVFQNLQVIRGRILHNGANSLTLOGLGISHLGSLRELSGLALIHNTHLCPHTV 480
Db 421 DLSVFQNLQVIRGRILHNGANSLTLOGLGISHLGSLRELSGLALIHNTHLCPHTV 480
QY 481 PWDQLFRRPHQALLHTANRPEDECVGEGALCHQACAGHCMPGPTCCVACSOFLRQEC 540
Db 481 PWDQLFRRPHQALLHTANRPEDECVGEGALCHQACAGHCMPGPTCCVACSOFLRQEC 540
QY 541 VEECVVLQGLPREYVNAHCLPCHBECQPNQSVTCGPEADQCVCARHXDPFCVAC 600
Db 541 VEECVVLQGLPREYVNAHCLPCHBECQPNQSVTCGPEADQCVCARHXDPFCVAC 600
QY 601 PSQVPRDLSYMPIMKFPDEBAGACPCPINCHSCVDLDDGCPABEQASPSTSVSAVWG 660
Db 601 PSQVPRDLSYMPIMKFPDEBAGACPCPINCHSCVDLDDGCPABEQASPSTSVSAVWG 660
QY 661 ILVVVLGVVFGLIKRRQOKIRYTRRLLQETLEVEPLTPSGAMENQAOMRILKETEL 720
Db 661 ILVVVLGVVFGLIKRRQOKIRYTRRLLQETLEVEPLTPSGAMENQAOMRILKETEL 720
QY 721 RKVKYLSGAGFNNFTYSFWLRP-----KVASHLEVLARENTPRANKETLIDEAVMA 775
Db 721 RKVKYLSGAGF---TVYKGIWIPDENYKIPVA-IKVARENTPRANKETLIDEAVMA 775
QY 776 GVGSPPYVSRLLGICLSTVQVLTQLMPEYGLLDHVRNRRGLSQDILLNMQOIAKMSY 835
Db 776 GVGSPPYVSRLLGICLSTVQVLTQLMPEYGLLDHVRNRRGLSQDILLNMQOIAKMSY 835
QY 836 LEDVRLVHRDLAARVLYKSNHYKITDFGLARLLDDEFBYHADGKVPKYMALLESIL 895
Db 836 LEDVRLVHRDLAARVLYKSNHYKITDFGLARLLDDEFBYHADGKVPKYMALLESIL 895
QY 896 RRRFTHQSDVMSYGTWVMTLTFGAKPFDGIPABEIPDLLEKSGRLPQPICTIDVYIM 955
Db 896 RRRFTHQSDVMSYGTWVMTLTFGAKPFDGIPABEIPDLLEKSGRLPQPICTIDVYIM 955
QY 956 VKCWMIDSECRPRFRELVSERMAADPQRFVVIQNEIDLGPASLSDSTFYRSLIEDDMG 1015
Db 956 VKCWMIDSECRPRFRELVSERMAADPQRFVVIQNEIDLGPASLSDSTFYRSLIEDDMG 1015
QY 1016 DLVDAEYLVPQGGFFCPDPAAPGAGWVHHRHSSSTRSGGDLTLGLPSEEBEAPRPL 1075
Db 1016 DLVDAEYLVPQGGFFCPDPAAPGAGWVHHRHSSSTRSGGDLTLGLPSEEBEAPRPL 1075
QY 1076 APSEGAGSDVDFDGDIGMAAKGLQSLPTHDPSPICRYSSEDPYVLPBETGGYAAPLTCSP 1135
Db 1076 APSEGAGSDVDFDGDIGMAAKGLQSLPTHDPSPICRYSSEDPYVLPBETGGYAAPLTCSP 1135
QY 1136 QPEYVNOQDVARPQPPSPREGPLPAARFAGATLIERPKTISPQKGVVDVAFGAVENPE 1195

|||||
Db 1136 QPEYVQPDVARPPSPSPREGPLPAPAPGATLERPKLSPKNGVVDVAFGAVENPE 1195
QY 1136 YLTPOGGAAPQHPPPAPSPAFNDLYVWDODPPRRGAPSPSPFKCTPAENPEYLIGLDVPP 1255
Db 1136 YLTPOGGAAPQHPPPAPSPAFNDLYVWDODPPRRGAPSPSPFKCTPAENPEYLIGLDVPP 1255
RESULT 12
ID AAE20479 standard; Protein; 1255 AA.
XX AAE20479;
XX AAE20479;
XX 01-JUL-2002 (first entry)
XX Human Her-2/neu protein.
XX Human Her-2/neu protein.
XX Human, Her-2/Neu protein; immune response; gene therapy; breast cancer;
XX human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX Homo sapiens.
XX Key Location/Qualifiers
XX Region 1021..1030
XX /note= "Naturally processed HLA-B44-restricted epitope"
XX WO200214503-A2.
XX 21-FEB-2002.
XX 14-AUG-2001; 2001MO-US41733.
XX 14-AUG-2000; 2000US-225152P.
XX 28-SEP-2000; 2000US-236428P.
XX 21-FEB-2001; 2001US-270520P.
XX (CORI-) CORIXA CORP.
XX Hand-Zimmermann S, Cheever MA, Foy TM, Lodee MJ, Kalos MD;
XX McNeil PD, Vedvick TS;
XX WPI; 2002-280758/32.
XX DR N-PSDB; AAD32743.
XX PT Novel isolated Her-2/Neu polypeptide composition useful for therapy,
XX prevention and diagnosis of cancer, preferably breast cancer
XX Disclosure; Page 114-117; 129pp; English.
XX The invention relates to an isolated Her-2/Neu polypeptide composition
XX effective for eliciting an immune response. The invention is useful for
XX eliciting an immune response in a patient, where the patient is human
XX leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
XX The composition is useful for the therapy and diagnosis of cancer.
XX preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
XX and other compositions for the diagnosis, prevention and treatment of
XX human malignancies, for stimulating and/or expanding T cells specific for
XX Her-2/Neu polypeptide and for inhibiting the development of cancer in a
XX patient. The invention is useful for stimulating a T cell response in a
XX human patient, as probe or primer for nucleic acid hybridisation, to
XX selectively form duplex molecules with complementary stretches of the
XX entire Her-2/Neu gene or gene fragments of interest, to isolate a full
XX length gene from a suitable library, and to direct expression of a
XX polypeptide in appropriate host cells. The composition is useful in
XX prophylactic or therapeutic applications and for the treatment of cancer,
XX preferably for the immunotherapy of breast cancer and other Her-2/Neu-
XX associated malignancies. The invention is useful in gene therapy. The
XX present sequence is human Her-2/neu protein.
SQ Sequence 1255 AA;
Query Match 96.9%; Score 6602; DB 23; Length 1255;

Best Local Similarity 96.9%; Pred. No. 0;
Matches 1221; Conservative 11; Mismatches 18; Indels 10; Gaps 3;
QY 1 MEIAALCRWGLLLALDPGAASTQVCTGTMTKRLPASPETHLDMKRLHYLQGGCVVQGNL 60
Db 1 MELAALCRWGLLLALDPGAASTQVCTGTMTKRLPASPETHLDMKRLHYLQGGCVVQGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVGGVYLLAHNOVRQVPLQRLRIVRGTOLFEDNVALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVGGVYLLAHNOVRQVPLQRLRIVRGTOLFEDNVALAVLDNG 120
QY 121 DPLNNTTPTVTSAPGRLRELQRLSLTEILKGVYLIGNPQLCYQDTILMDIFHKNNQLA 180
Db 121 DPLNNTTPTVTSAPGRLRELQRLSLTEILKGVYLIGNPQLCYQDTILMDIFHKNNQLA 180
QY 181 LTLIDTNSRACHCSPCKSGSRCKWGSSESDCGLTRTVACGACGACGKPLPTDCHEQC 240
Db 181 LTLIDTNSRACHCSPCKSGSRCKWGSSESDCGLTRTVACGACGACGKPLPTDCHEQC 240
QY 241 AAGCTGPRHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPRHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVSGCTLVCPPLHNOEVTAECDTORCEKSKRCARVCGAGQYIKANSKFGIT 360
Db 301 YNYLSTDVSGCTLVCPPLHNOEVTAECDTORCEKSKRCARVCGAGQYIKANSKFGIT 360
QY 361 ELEFAGCKKIFGSLAFIPESFDGPASNTAPLQEQVETLEITGYLISAMPDSLP 420
Db 361 IOEFAGCKKIFGSLAFIPESFDGPASNTAPLQEQVETLEITGYLISAMPDSLP 420
QY 421 DLSVFQNLQVIRGRILNHGAYSLTLOGISISULGRLBELGSLALIHNTHLCPHTV 480
Db 421 DLSVFQNLQVIRGRILNHGAYSLTLOGISISULGRLBELGSLALIHNTHLCPHTV 480
QY 481 PWDOLFRRPHQALHTANRPDECEVSGGLACHQACGRCMPPTCCVACSGFLRQEC 540
Db 481 PWDOLFRRPHQALHTANRPDECEVSGGLACHQACGRCMPPTCCVACSGFLRQEC 540
QY 541 VEECRVLQGLFREYVNAHCLPCHPECOQNGSVTCGPADQCVCACAHKDPFCVARC 600
Db 541 VEECRVLQGLFREYVNAHCLPCHPECOQNGSVTCGPADQCVCACAHKDPFCVARC 600
QY 601 PSQVPRDLSVMPIMKPPREBGAQCPICINCHSCVDLDDGCPAEQASLTSIVSVVG 660
Db 601 PSQVPRDLSVMPIMKPPREBGAQCPICINCHSCVDLDDGCPAEQASLTSIVSVVG 660
QY 661 ILVVVLGVVFGILIKRQOKIRKRYTRRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILVVVLGVVFGILIKRQOKIRKRYTRRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 721 RKVYLGSGAGFNNTYSEFWLRP-----KVSASHLEVRRENTSPKANEILDEAVYMA 775
Db 721 RKVYLGSGAGFNNTYSEFWLRP-----KVSASHLEVRRENTSPKANEILDEAVYMA 775
QY 776 GVGSYVYRLLIGICLTSTVOLVOTLMPEYGCILDHVRNRRGLSODILNMCQIAKMSY 835
Db 776 GVGSYVYRLLIGICLTSTVOLVOTLMPEYGCILDHVRNRRGLSODILNMCQIAKMSY 835
QY 836 LEDVALVRDLAARNVLYKSPNHYKITDFIGLRLDIDETRYHADGGKVIDKMALESIL 895
Db 836 LEDVALVRDLAARNVLYKSPNHYKITDFIGLRLDIDETRYHADGGKVIDKMALESIL 895
QY 896 RRRPFDHSDVSYGVATWELMTFGAKYDGIIPAREIDLEKEBRLPOPICITDYYMIM 955
Db 896 RRRPFDHSDVSYGVATWELMTFGAKYDGIIPAREIDLEKEBRLPOPICITDYYMIM 955
QY 956 VKCWMIDSECRPRRELIVSEFSRMAQDPQFFVYIQNEDLGPASPLDSTFYRLSLEDDEMG 1015
Db 956 VKCWMIDSECRPRRELIVSEFSRMAQDPQFFVYIQNEDLGPASPLDSTFYRLSLEDDEMG 1015
QY 1016 DLVDAEYLVPOGFPCEPDAPAPAGWGHRRHSSSTRSGGGLTGLGESSEEARSPSL 1075
Db 1016 DLVDAEYLVPOGFPCEPDAPAPAGWGHRRHSSSTRSGGGLTGLGESSEEARSPSL 1075

Db 1016 DLVDAEYIVPQGGFCPPDPAGAGWYHHRSSSTRSGGDLTLGLSPSEBAPRSP 1075
QY 1076 APSEGAGSDVPDGLGMGAAGKGLSLPHDPSPLQRYSEDPVPLPSETDGYAPLTCS 1135
Db 1076 APSEGAGSDVPDGLGMGAAGKGLSLPHDPSPLQRYSEDPVPLPSETDGYAPLTCS 1135
QY 1136 OPEYVNPQVRPQPSPPRGPLPARPPGATLERKTLSPKNGVYMDVFAFGAVENPE 1195
Db 1136 OPEYVNPQVRPQPSPPRGPLPARPPGATLERKTLSPKNGVYMDVFAFGAVENPE 1195
QY 1196 YLTPOGGAAPQHPAPSPAPFNDLVYMDQDPERKAPSTFKGTANENPEYGLDVPV 1255
Db 1196 YLTPOGGAAPQHPAPSPAPFNDLVYMDQDPERKAPSTFKGTANENPEYGLDVPV 1255
RESULT 13
AAM51143
ID AAM51143 standard; Protein; 1255 AA.
XX
AC AAM51143;
XX
DT 17-JUN-2002 (first entry)
XX
DE Human Her-2/neu oncogene-encoded p185 glycoprotein.
XX
KM Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
XX tyrosine kinase; receptor; C-erbB2; gene therapy.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Domain 1..653 /note="extracellular domain"
FT Domain 676..1255 /note="intracellular domain"
FT Domain 990..1255 /note="phosphorylation domain"
FT
PN WO200212341-A2.
XX
PD 14-FEB-2002.
XX
PF 03-AUG-2001; 2001WO-US24283.
XX
PR 03-AUG-2000; 2000US-0632507.
XX
PA (CORI-) CORIXA CORP.
XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
PI Cheever MA, Gheysen D;
XX
XX MPI; 2002-241743/29.
DR N-PSDB; ABA92250.
XX
PT Her-2/neu fusion protein for treating or preventing cancer by eliciting
PT extracellular domain fused to Her-2/neu intracellular or
PT phosphorylation domain -
XX
PS Claim 68; Fig 7; 14pp; English.
XX
XX The present sequence is that of human Her-2/neu (p185 glycoprotein
CC or C-erbB2), an oncogenic self-protein and target for anti-cancer
CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
CC in a variety of cancers, including breast, ovarian, colon, lung and
CC prostate cancer. Her-2/neu is a member of the tyrosine kinase
CC family of receptor-like glycoproteins. It comprises an extracellular
CC domain with homology to the epidermal growth factor receptor
CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
CC intracellular domain that also shows homology to EGFR. Its
CC overexpression correlates with a poor prognosis in breast and
CC ovarian cancers. The invention provides Her-2/neu fusion
CC proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
CC preferred fusion proteins, the extracellular domain of a Her-2/neu
CC protein is fused to a Her-2/neu intracellular domain or
CC phosphorylation domain (or its deleted fragment). An immune
CC response to Her-2/neu protein is elicited or enhanced by
CC administering the fusion protein in the form of a vaccine, or by
CC transfecting cells of an animal *ex vivo* with a nucleic acid
CC encoding the fusion protein, and delivering the transfected cells
CC to the animal. The fusion proteins, nucleic acids, and isolated
CC specific T-cells are useful for inhibiting the development of a
CC cancer, especially breast, ovarian, colon, lung or prostate cancer
CC in a patient. T cells that specifically react with a Her-2/neu
CC fusion protein can be used to remove tumour cells from a sample in
CC order to inhibit the development of cancer in a patient.
XX
SQ Sequence 1255 AA;
Query Match 96.9%; Score 6602; DB 23; Length 1255;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 1221; Conservative 11; Mismatches 18; Indels 10; Gaps 3;
QY 1 MELALCWGILLALPPGAASVQVCTGDKRLPASPETHLMDLHLYOGCVQGNL 60
Db 1 MELALCWGILLALPPGAASVQVCTGDKRLPASPETHLMDLHLYOGCVQGNL 60
QY 61 ELTYLPTNASLFLQDIOEVQGYVLIANQVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLFLQDIOEVQGYVLIANQVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTGASPGGLRELDQRLSITELKGGVLIQRPOLCTODTLMKDIFPKNNQLA 180
Db 121 DPLNNTTPTGASPGGLRELDQRLSITELKGGVLIQRPOLCTODTLMKDIFPKNNQLA 180
QY 181 LTLIDTNSRACHPSPCKGSRCKGSESEDCQSLTRVCGGACGARKGRLPTCCHEQC 240
Db 181 LTLIDTNSRACHPSPCKGSRCKGSESEDCQSLTRVCGGACGARKGRLPTCCHEQC 240
QY 241 AAGCTGPKHSDCLAFHNHSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTACP 300
Db 241 AAGCTGPKHSDCLAFHNHSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTACP 300
QY 301 YNYLSTDVSGCTLVGPLELNOEVTLEDGQRECKSKPARVCYGLGMVYIANKKFGIT 360
Db 301 YNYLSTDVSGCTLVGPLELNOEVTLEDGQRECKSKPARVCYGLGMVYIANKKFGIT 360
QY 361 ELEFAGCKKIFGSLAFLPESFPDGPASNTAPLOEOLQVFETLEITGYLYISAMPDLP 420
Db 361 ELEFAGCKKIFGSLAFLPESFPDGPASNTAPLOEOLQVFETLEITGYLYISAMPDLP 420
QY 421 DLSVFQNLQVIRGRIILHNGAYSILTQGLGISMGLRSIRELGGSLALIHNTHLCPVHTV 480
Db 421 DLSVFQNLQVIRGRIILHNGAYSILTQGLGISMGLRSIRELGGSLALIHNTHLCPVHTV 480
QY 481 PMDQLFRRNHQALLTANRPEDECGEGSLAQHOLCARHCGMPPTQCVCNCSQFLRQEC 540
Db 481 PMDQLFRRNHQALLTANRPEDECGEGSLAQHOLCARHCGMPPTQCVCNCSQFLRQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECOPQNGSVTCFGEADQCAAHYKDPFVCARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECOPQNGSVTCFGEADQCAAHYKDPFVCARC 600
QY 601 PSQVNPDSLYMIMKFPBEGACQCPINCHSCYDLDKCCPMEORASPLTISAVVG 660
Db 601 PSQVNPDSLYMIMKFPBEGACQCPINCHSCYDLDKCCPMEORASPLTISAVVG 660
QY 661 ILLVYVLGVVFILIKRQOKIRKXTMRLLQETELVPLTPSGAMPQAOMLIKETEL 720
Db 661 ILLVYVLGVVFILIKRQOKIRKXTMRLLQETELVPLTPSGAMPQAOMLIKETEL 720
QY 721 RKVKVLGSAFENFTYSPMLRVF----KVSASHLEVLRENTSPRANKETLDEAVYMA 775
Db 721 RKVKVLGSAFENFTYSPMLRVF----KVSASHLEVLRENTSPRANKETLDEAVYMA 775

QY 776 GVGSFVYSRLIGICLTSTVQVLTQMPYGCILDHVRENRGRIGSGDILNMCQIAKMSY 835
DB 776 GVGSFVYSRLIGICLTSTVQVLTQMPYGCILDHVRENRGRIGSGDILNMCQIAKMSY 835
QY 836 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMALLESIL 895
DB 836 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMALLESIL 895
QY 896 RRRFTHQSDVWSYGVYVWELMTFGAKFYDGIIPAREIPDLEKGERLPQPICTIDVYIM 955
DB 896 RRRFTHQSDVWSYGVYVWELMTFGAKFYDGIIPAREIPDLEKGERLPQPICTIDVYIM 955
QY 956 VKCMIMIDSECRPFRELVSEFSRMAPDPQRFVYIQUEDLGASPLDSTFFYSLLLEDDMG 1015
DB 956 VKCMIMIDSECRPFRELVSEFSRMAPDPQRFVYIQUEDLGASPLDSTFFYSLLLEDDMG 1015
QY 1016 DLVDAEEYLVPOGFFCDDPAPAGAGMVRHRSSTSGGSDLTLGEPSEEAAPSPL 1075
DB 1016 DLVDAEEYLVPOGFFCDDPAPAGAGMVRHRSSTSGGSDLTLGEPSEEAAPSPL 1075
QY 1076 APSEGAGSDVFDGDLGMAKGLQSLPTHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSP 1135
DB 1076 APSEGAGSDVFDGDLGMAKGLQSLPTHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSP 1135
QY 1136 QPEYVNOQDVPRQPPSPREGRPLPAPAPAGATLERAKTILSPGNKVYKQVFAFGAVENPE 1195
DB 1136 QPEYVNOQDVPRQPPSPREGRPLPAPAPAGATLERAKTILSPGNKVYKQVFAFGAVENPE 1195
QY 1196 YLTPQGGAAPQHPPPAPFPAFDNLVYWDQDPERGAPSTFKGTPTAENPEYGLDVPV 1255
DB 1196 YLTPQGGAAPQHPPPAPFPAFDNLVYWDQDPERGAPSTFKGTPTAENPEYGLDVPV 1255

RESULT 14
AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX AAU77114;
XX
XX
XX 05-JUN-2002 (first entry)
XX
XX Human Her-2/neu polypeptide.
XX
XX Human; Her-2/neu; leukaemia; haematological malignancy; CML;
XX acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
XX chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
XX Hodgkin's lymphoma; T cell therapy.
XX
XX Homo sapiens.
XX
XX MO200213847-A2.
XX
XX 21-FEB-2002.
XX
XX 13-AUG-2001; 2001MO-US25408.
XX
XX 14-AUG-2000; 2000US-0638280.
XX 28-SEP-2000; 2000US-0675904.
XX
XX (CORI-) CORIXA CORP.
XX
XX Gaiger A, Cheever MA, Hand-zimmermann S;
XX MPI; 2002-280741/32.
XX N-PSDB; ABK10730.
XX
XX Inhibiting haematological malignancy development by administering
XX polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
XX encoding the polypeptide, or antigen presenting cells expressing the
XX polypeptide
XX
XX Disclosure; Page 71-74; 74pp; English.
XX

CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myeloma, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;

Query Match 96.9%; Score 6602; DB 23; Length 1255;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 1221; Conservative 11; Mismatches 18; Indels 10; Gaps 3;

QY 1 MELAALCRWGLLALLPFGAASVTCVCTGDKRLPASPETHLDMRLVYGCQVQCNL 60
DB 1 MELAALCRWGLLALLPFGAASVTCVCTGDKRLPASPETHLDMRLVYGCQVQCNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQYVLIANQVROVPLQRLRIVRGTOLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIOEVQYVLIANQVROVPLQRLRIVRGTOLFEDNYALAVLDNG 120
QY 121 DPLANTPTTGASPGGLRBLRLSTILKSGVLIQRNPOLCYQDTILMKDIFHKNOILA 180
DB 121 DPLANTPTTGASPGGLRBLRLSTILKSGVLIQRNPOLCYQDTILMKDIFHKNOILA 180
QY 181 LTLIDPNSRACHPSPCKSGRCMGSESEDCQSITRVCAAGGACRCKGPIPTDCHEQC 240
DB 181 LTLIDPNSRACHPSPCKSGRCMGSESEDCQSITRVCAAGGACRCKGPIPTDCHEQC 240
QY 241 AAGCTGPKASDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGAYTGASCVTACP 300
DB 241 AAGCTGPKASDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGAYTGASCVTACP 300
QY 301 YNVLSTDVSGCTLVCPLNQEVYTAEDGQRECKSKPCARVCYGLGMQYIYANSKFIGIT 360
DB 301 YNVLSTDVSGCTLVCPLNQEVYTAEDGQRECKSKPCARVCYGLGMQYIYANSKFIGIT 360
QY 361 ELEFAGCKKIFGSLAFLPESFPDQASNTAFLQPEOLQVFETLEITGYLYISAMPDLSL 420
DB 361 IOEPAGCKKIFGSLAFLPESFPDQASNTAFLQPEOLQVFETLEITGYLYISAMPDLSL 420
QY 421 DLSVQNLQVIRGLIHNAGASLTIGIGISWLGIRSLREISGLALIHNTHLCEVHTV 480
DB 421 DLSVQNLQVIRGLIHNAGASLTIGIGISWLGIRSLREISGLALIHNTHLCEVHTV 480
QY 481 PWDOLFERNHQLHTANRPEDECVGEGACIHOUCARHGCMGPQTCVNSOFLRGQEC 540
DB 481 PWDOLFERNHQLHTANRPEDECVGEGACIHOUCARHGCMGPQTCVNSOFLRGQEC 540
QY 541 VEECRVLOGLPREYVNAHCLPCHBECOPQNGSVTCFGEADQCAACHYXDPFCVARC 600
DB 541 VEECRVLOGLPREYVNAHCLPCHBECOPQNGSVTCFGEADQCAACHYXDPFCVARC 600
QY 601 PSGVPRDLSYMPIMKFPDEEGACQPCINCHSCVDLDKCPAQRSPILTSIISAVG 660
DB 601 PSGVPRDLSYMPIMKFPDEEGACQPCINCHSCVDLDKCPAQRSPILTSIISAVG 660
QY 661 ILLVVLGVVFGILIKRQOKIRKTYRRLLOETELVEPLTPSGAMPQAOMLTKETEL 720
DB 661 ILLVVLGVVFGILIKRQOKIRKTYRRLLOETELVEPLTPSGAMPQAOMLTKETEL 720
QY 721 RKVKYLGSGAFGFPNNFTYSFWLRVP-----KVSASHLEVLRENTSPRANKELDEAYYMA 775
DB 721 RKVKYLGSGAFG---TYKGIWIPDGENVKIPVA-IKVARENTSPRANKELDEAYYMA 775
QY 776 GVGSFVYSRLIGICLTSTVQVLTQMPYGCILDHVRENRGRIGSGDILNMCQIAKMSY 835
DB 776 GVGSFVYSRLIGICLTSTVQVLTQMPYGCILDHVRENRGRIGSGDILNMCQIAKMSY 835
QY 836 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMALLESIL 895
DB 836 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMALLESIL 895

```

Db      |||||
636 LEDVRLVHRDLAARNLVKSPNHVKITDGLARLLDIDETEHADGKVP1KMALESIL 895
QY      RRRFTHQSDVWSYGVTVWELMTFGAKPYDGI.PAREIPDLLEKGERLPQPICTIDVYIM 955
Db      896 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGI.PAREIPDLLEKGERLPQPICTIDVYIM 955
QY      956 VKCWMIDSECRPRFRELVSFRRMARDPQRFVVIQNEDLGPASPLDSTFYSLLEDDMG 1015
Db      956 VKCWMIDSECRPRFRELVSFRRMARDPQRFVVIQNEDLGPASPLDSTFYSLLEDDMG 1015
QY      1016 DLVDAEEYLVPOQGFCCPDPAAGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRSPL 1075
Db      1016 DLVDAEEYLVPOQGFCCPDPAAGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRSPL 1075
QY      1076 APSEGAGSDVFDGDLGMGAAGLQSLPTHDPSPLOKYSBDPTVPLPSTDGVVAPLTGSP 1135
Db      1076 APSEGAGSDVFDGDLGMGAAGLQSLPTHDPSPLOKYSBDPTVPLPSTDGVVAPLTGSP 1135
QY      1136 QPEYVQOPDVPRQPPSPREGPLPAARPAATLERAKTILSPKNGVYKDVAFAGAVENPE 1195
Db      1136 QPEYVQOPDVPRQPPSPREGPLPAARPAATLERAKTILSPKNGVYKDVAFAGAVENPE 1195
QY      1196 YLTPQGAAPQHPHPAFSPAFDNLTYWDQDPPERGAPSTFKGTPTAENPEYGLDVPV 1255
Db      1196 YLTPQGAAPQHPHPAFSPAFDNLTYWDQDPPERGAPSTFKGTPTAENPEYGLDVPV 1255

RESULT 15
AAR39568
ID AAR39568 standard; Protein; 1433 AA.
AC AAR39568;
XX
DT 07-FEB-1994 (first entry)
DE Sequence of c-erbB-2 tumour antigen.
KM Tumour antigen; c-erbB-2; glycoprotein.
XX
OS Homo sapiens.
PN M09316185-A.
XX
PD 19-AUG-1993.
XX
PF 05-FEB-1993; 93MO-US01055.
XX
PR 06-FEB-1992; 92US-0831967.
XX
PA (CETU) CETUS ONCOLOGY CORP.
PA (CREA-) CREATIVE BIOMOLECULES INC.
PI Houston IL, Huston JS, Oppermann H, Ring DB;
XX
DR WPI: 1993-272889/34.
DR N-PSDB: AAQ46083.
XX
PT New single chain Fv polypeptide binding to C-erbB-2 tumour
PT antigen - for imaging or treating breast or ovarian cancer etc.
XX
PS Disclosure; pages 48-54; 87pp; English.
XX
CC c-erbB-2 refers to a protein antigen expressed on the surface of
CC tumour cells such as breast and ovarian tumour cells, which is an
CC apptox. 200,000 mol. wt. acidic glycoprotein having an isoelectric
CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents
CC the location of a stop codon in AAQ46083.
XX
SQ Sequence 1433 AA;
Query March 96.3%; Score 6559; DB 14; Length 1433;
Best Local Similarity 96.3%; Pred. No. 0;

```

```

Matches 1214; Conservative 13; Mismatches 23; Indels 10; Gaps 3;
QY      1 MELALCRWGLLALLPPGAASVQVCTGTDKMLPLPASPETHLDMLEHLVQGCQVQGNL 60
Db      1 MELALCRWGLLALLPPGAASVQVCTGTDKMLPLPASPETHLDMLEHLVQGCQVQGNL 60
QY      61 ELTYLPTNASLSFLDIOEVQGYVLIANQVRQVPLQRLIRVGTQCFEDYVALAVLDNG 120
Db      61 ELTYLPTNASLSFLDIOEVQGYVLIANQVRQVPLQRLIRVGTQCFEDYVALAVLDNG 120
QY      121 DPLNNTPTVPGASPGGLRELOJLSRTEILKGGVLIQRNPOLCYQDTILMWDIFKNNOLA 180
Db      121 DPLNNTPTVPGASPGGLRELOJLSRTEILKGGVLIQRNPOLCYQDTILMWDIFKNNOLA 180
QY      181 LTLIDTNSRACHPCSPCKSGRCMGSESSDQSLTRVACGAGCARCGPLPTCCHEQC 240
Db      181 LTLIDTNSRACHPCSPCKSGRCMGSESSDQSLTRVACGAGCARCGPLPTCCHEQC 240
QY      241 AAGCTGPKSDCLACLFHNSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVTACP 300
Db      241 AAGCTGPKSDCLACLFHNSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVTACP 300
QY      301 YNLTSDVSSCTLVCPILNQEVTADGTQREKSGKPCARVCYGLGMQYIKANSKFIGIT 360
Db      301 YNLTSDVSSCTLVCPILNQEVTADGTQREKSGKPCARVCYGLGMQYIKANSKFIGIT 360
QY      361 ELEFAGCKKIFGSLAFLESPFGDPAASNTAFLQEPOLQVFETLEITGYLYISAMPDLP 420
Db      361 IGFAPGCKKIFGSLAFLESPFGDPAASNTAFLQEPOLQVFETLEITGYLYISAMPDLP 420
QY      421 DLSVFQNLQVIRGRILHNGAVSLTQGGISMLGILSRLEISGIALIHNHTHLSFVHTV 480
Db      421 DLSVFQNLQVIRGRILHNGAVSLTQGGISMLGILSRLEISGIALIHNHTHLSFVHTV 480
QY      481 PMDQLEFRNHQALLHTANRPEDECEVGEGLACHQICARHGCMGPGTQCVCNCSQFLRGEC 540
Db      481 PMDQLEFRNHQALLHTANRPEDECEVGEGLACHQICARHGCMGPGTQCVCNCSQFLRGEC 540
QY      541 VBECRYLQGLPREYVNAARHCLPCHBECOPQNSVTCFGEADQCAACHYKDPFCVARC 600
Db      541 VBECRYLQGLPREYVNAARHCLPCHBECOPQNSVTCFGEADQCAACHYKDPFCVARC 600
QY      601 PSQVPRDLSYMEIWKPPDEBGAQCPCPINCHSCVDLDKCPAQRAPSLTISAVAG 660
Db      601 PSQVPRDLSYMEIWKPPDEBGAQCPCPINCHSCVDLDKCPAQRAPSLTISAVAG 660
QY      661 ILLVVLGVVFGILLIKRQCKIRKTYMRLQETELVPLTPSGAMPQAOARILKETEL 720
Db      661 ILLVVLGVVFGILLIKRQCKIRKTYMRLQETELVPLTPSGAMPQAOARILKETEL 720
QY      721 RKVKYVGSQAFGEFNFTUSFMLRVP----KVSASHLEVLRENTSPKANKELIDEAAYMA 775
Db      721 RKVKYVGSQAFGEFNFTUSFMLRVP----KVSASHLEVLRENTSPKANKELIDEAAYMA 775
QY      776 GVGSPIVSRLLGICLTSITQVLTQVLTQVLTQVLTQVLTQVLTQVLTQVLTQVLTQVLTQV 835
Db      776 GVGSPIVSRLLGICLTSITQVLTQVLTQVLTQVLTQVLTQVLTQVLTQVLTQVLTQV 835
QY      836 LEDVRLVHRDLAARNLVKSPNHVKITDGLARLLDIDETEHADGKVP1KMALESIL 895
Db      836 LEDVRLVHRDLAARNLVKSPNHVKITDGLARLLDIDETEHADGKVP1KMALESIL 895
QY      896 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGI.PAREIPDLLEKGERLPQPICTIDVYIM 955
Db      896 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGI.PAREIPDLLEKGERLPQPICTIDVYIM 955
QY      956 VKCWMIDSECRPRFRELVSFRRMARDPQRFVVIQNEDLGPASPLDSTFYSLLEDDMG 1015
Db      956 VKCWMIDSECRPRFRELVSFRRMARDPQRFVVIQNEDLGPASPLDSTFYSLLEDDMG 1015
QY      1016 DLVDAEEYLVPOQGFCCPDPAAGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRSPL 1075
Db      1016 DLVDAEEYLVPOQGFCCPDPAAGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRSPL 1075

```

Qy	1076	APSEGAGSVFENDIGMGAKGLOSLPTHDPSPLOQYSEDPVTPLPSETDGYVAPLTCSP	1135
Db	1076	APSEGAGSDVFDGDLGMGAKGLOSLPTHDPSPLOQYSEDPVTPLPSETDGYVAPLTCSP	1135
Qy	1136	QPEYVNOQDVPRQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAPFGAVENPE	1195
Db	1136	QPEYVNOQDVPRQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAPFGAVENPE	1195
Qy	1196	YLTPOGGAAPQPHPPAFSPAFDNLVYWDQDPPERGAPPSFTFKGTPTAENPEYLGLOVPV	1255
Db	1196	YLTPOGGAAPQPHPPAFSPAFDNLVYWDQDPPERGAPPSFTFKGTPTAENPEYLGLOVPV	1255

Search completed: July 22, 2003, 09:17:00
 Job time : 41.9319 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2793 Seconds

(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-325-339-14

Perfect score: 6852

Sequence: 1 MELALCRWGLLALPPGA.....TFKGTPTAENPSYGLDVPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

112892

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Result No. Score Match Length DB ID Description

Result No.	Score	Match	Length	DB	ID	Description
1	6730	98.2	1255	1	ERB2_HUMAN	P04626 homo sapien
2	5931	86.6	1257	1	ERB2_RAT	P06494 ratius norv
3	5916.5	86.3	1254	1	ERB2_MESAU	Q60553 mesocricetu
4	3148	45.9	1210	1	EGFR_HUMAN	P00533 homo sapien
5	3124	45.6	1210	1	EGFR_MOUSE	001279 mus musculu
6	2981.5	43.5	1308	1	ERB4_HUMAN	015303 homo sapien
7	2862	43.2	1308	1	ERB4_RAT	062966 ratius norv
8	2688.5	39.2	1167	1	XMRK_XIPMA	P13388 xiphophorus
9	2439.5	35.6	1342	1	ERB3_HUMAN	P21860 homo sapien
10	2365.5	34.5	1339	1	EGFR_DROME	P04412 dirosophila
11	1965	28.7	1426	1	EGFR_DROME	P04412 dirosophila
12	1749.5	25.5	634	1	ERBB_AVIER	P00534 avian leuko
13	1703	24.9	604	1	ERBB_AVIER	P00535 avian eryth
14	1630	23.8	540	1	ERBB_AVIER	P11273 avian eryth
15	1600	23.4	703	1	EGFR_CHICK	P13387 gallus gall
16	1300	19.0	1323	1	LT23_CABEL	P24348 caenorhabdi
17	1142.5	16.7	245	1	ERB2_MOUSE	P70424 mus musculu
18	730	10.7	1363	1	ILPR_BRALA	002466 brachiofoco
19	696	10.2	1382	1	INSR_HUMAN	P06213 homo sapien
20	693	10.1	1300	1	IRP_MOUSE	09w14 mus musculu
21	693	10.1	1607	1	MLP_LYMS	Q25410 lymnaea sta
22	691	10.1	1383	1	INSR_RAT	P15147 ratius norv
23	690.5	10.1	1372	1	INSR_MOUSE	P15208 mus musculu
24	689	10.1	1477	1	HTK7_HYPAT	Q25197 hydra atten
25	684	10.0	1297	1	IRR_HUMAN	P14616 homo sapien
26	679.5	9.9	1300	1	IRP_CAVPO	P14617 cavia porce
27	638	9.1	1367	1	IGIR_MOUSE	P08069 homo sapien
28	622	9.1	1373	1	IGIR_MOUSE	P08071 mus musculu
29	621	9.1	2146	1	INSR_DROME	P09208 dirosophila
30	618.5	9.0	1370	1	IGIR_RAT	P24062 ratius norv
31	617	9.0	1390	1	INSR_AEDAE	Q93105 aedes aegypt
32	602	8.8	987	1	ERB4_HUMAN	P54760 homo sapien
33	598.5	8.7	984	1	ERB1_CHICK	Q07494 gallus gall

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

34	596	8.7	976	1	EPB2_HUMAN	P29317 homo sapien
35	590.5	8.6	977	1	EPB2_MOUSE	Q03145 mus musculu
36	588	8.6	902	1	EPBB_XENTIA	Q91736 xenopus lae
37	588	8.6	987	1	EPB4_MOUSE	P54761 mus musculu
38	588	8.5	1114	1	RET_HUMAN	P07949 homo sapien
39	583.5	8.5	984	1	EPB1_RAT	P09759 ratius norv
40	582.5	8.5	985	1	EPBA_XENTIA	Q91571 xenopus lae
41	577.5	8.4	984	1	EPB1_HUMAN	P54762 homo sapien
42	577	8.4	757	1	HT16_HYPAT	P53356 hydra atten
43	573.5	8.4	1053	1	FAK1_CHICK	Q00944 gallus gall
44	569	8.3	1068	1	FAK1_XENTIA	Q91738 xenopus lae
45	563	8.2	1052	1	FAK1_MOUSE	P34152 mus musculu

ALIGNMENTS

RESULT 1
ID ERB2_HUMAN STANDARD; PRT; 1255 AA.
AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor Protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-cyte cell
DE surface receptor HER2) (MLN 19).
GN ERB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=66118663; PubMed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA Saito T., Toyoshima K.;
RT "Similarity of protein encoded by the human c-erbB-2 gene to
RT epidermal growth factor receptor.";
RL Nature 319:230-234(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86070181; PubMed=2999974;
RA Cousens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA McGarrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,
RA Francke U., Levinson A., Ulrich A.;
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene.";
RL Science 230:1132-1139(1985).
RN [3]
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=86016725; PubMed=2995967;
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.;
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor receptor gene and is amplified in a
RT human salivary gland adenocarcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN [4]
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT "Characterization of a new allele of the human ERBB2 gene by allele-
RT specific competition hybridization.";
RL Genomics 15:426-429(1993).
RN [5]
RP FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
RN ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
RN POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
RN ALPHA AND AMPHIREGULIN.
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -I- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIALLY).
CC -I- SUBCELLULAR LOCATION: Type I membrane protein.

FT DISULFID 517 530 BY SIMILARITY.
 FT DISULFID 533 542 BY SIMILARITY.
 FT DISULFID 546 562 BY SIMILARITY.
 FT DISULFID 565 578 BY SIMILARITY.
 FT DISULFID 569 586 BY SIMILARITY.
 FT DISULFID 589 598 BY SIMILARITY.
 FT DISULFID 602 625 BY SIMILARITY.
 FT DISULFID 628 636 BY SIMILARITY.
 FT DISULFID 632 644 BY SIMILARITY.
 FT MOD.RES 1141 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD.RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 188 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
 SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 86.6%; Score 5931; DB 1; Length 1257;
 Best Local Similarity 86.7%; Pred. No. 1,2e-309;
 Matches 1096; Conservative 56; Mismatches 102; Indels 10; Gaps 4;

QY 1 MELALALCRWGLLALBPAGASTQVCTGTDMLKLPASPEHLDMLRHLYQGCGVYQGNL 60
 DB 1 MELAMCRWGLLALBPAGIAGTQCTGTDMLKLPASPEHLDMLRHLYQGCGVYQGNL 60
 QY 61 ELTYLPTNASTLFLODIOEVGYLIANQVRCQVPLQRLRIIVEGTQLFEDNVALAVLDNG 120
 DB 61 ELTYLPANASTLFLODIOEVGYLIANQVRCQVPLQRLRIIVEGTQLFEDNVALAVLDNR 120
 QY 121 DPLNNTTPTVT-GASFGGLRELIQLRSLTEILKGVLIQRNPQLCYQDTILMKDI-FHKNQNL 179
 DB 121 DPQDNVAASTPGRTPREGRLRELQRLSTLEILKGVLIQRNPQLCYQDNVLMKDVFRKXNQL 180
 QY 180 ALTLIDITNRSRACHCSPMKGRCWGSSEDCOSLRITVCAAGCAKCPRLPTDCHEG 239
 DB 181 APVDIDITNRSRACHCSPMKGRCWGSSEDCOSLRITVCAAGCAKCPRLPTDCHEG 240
 QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNPREGRTFGASCCTAC 299
 DB 241 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNPREGRTFGASCCTAC 300
 QY 300 PYNVLTVDGSGCTVACPLHNOEVTAEQTCRCCKSPCARVCYGLKMFNFVTSWLN 359
 DB 301 PYNVLTVDGSGCTVACPLHNOEVTAEQTCRCCKSPCARVCYGLKMFNFVTSWLN 359
 QY 360 PK-VASHLSEFPAGCKIFGSLAPLPESFDQDPASNTAPLOPEQLQVFTLEIEITGYLYT 418
 DB 360 PK-VASHLSEFPAGCKIFGSLAPLPESFDQDPASNTAPLOPEQLQVFTLEIEITGYLYT 418
 QY 419 SAMPDSLPLDLSYFQNLQVIRGRILHNGVASTLQGLGISMGLRSLAEISGLALIHANT 478
 DB 419 SAMPDSLPLDLSYFQNLQVIRGRILHNGVASTLQGLGISMGLRSLAEISGLALIHANT 478
 QY 414 SAMPDSLPLDLSYFQNLQVIRGRILHNGVASTLQGLGISMGLRSLAEISGLALIHANT 473
 DB 414 SAMPDSLPLDLSYFQNLQVIRGRILHNGVASTLQGLGISMGLRSLAEISGLALIHANT 473
 QY 479 HLCFVHTVPMQDLFENPHQALLHTANPEDE-CVGEGLACHQOLCARHCGMPPTQCVNC 537
 DB 479 HLCFVHTVPMQDLFENPHQALLHTANPEDE-CVGEGLACHQOLCARHCGMPPTQCVNC 537
 QY 538 SGLFGRQCEVEBCRYLQGLPREYVNAHCHLCPHECOPQNGSVTCFGEPAEDQVACAHYK 597
 DB 538 SGLFGRQCEVEBCRYLQGLPREYVNAHCHLCPHECOPQNGSVTCFGEPAEDQVACAHYK 597
 QY 534 SHFLRGQCEVEBCRYLQGLPREYVNAHCHLCPHECOPQNGSVTCFGEPAEDQVACAHYK 593
 DB 534 SHFLRGQCEVEBCRYLQGLPREYVNAHCHLCPHECOPQNGSVTCFGEPAEDQVACAHYK 593
 QY 598 DPPFCVACPSGVKXDLSTYPMIWKFPDEGACOPCINCHSCVLDLDDKCPMEGRASPL 657
 DB 598 DPPFCVACPSGVKXDLSTYPMIWKFPDEGACOPCINCHSCVLDLDDKCPMEGRASPL 657
 QY 594 DSSSCVACPSGVKXDLSTYPMIWKFPDEGACOPCINCHSCVLDLDDKCPMEGRASPL 653
 DB 594 DSSSCVACPSGVKXDLSTYPMIWKFPDEGACOPCINCHSCVLDLDDKCPMEGRASPL 653
 QY 658 TSIVAVVGLILVVLGVFGLILKRRQCKIKYTRMLLOETELVPLTPSGAMPVQAQ 717
 DB 658 TSIVAVVGLILVVLGVFGLILKRRQCKIKYTRMLLOETELVPLTPSGAMPVQAQ 717
 QY 654 TFIITAVGVLLFLILVYVGLILKRRQCKIKYTRMLLOETELVPLTPSGAMPVQAQ 713
 DB 654 TFIITAVGVLLFLILVYVGLILKRRQCKIKYTRMLLOETELVPLTPSGAMPVQAQ 713
 QY 718 MRLKETELRKVKVLGSAFGTVYKGIIPDGENVKIPVAIKVIRENTSPKANKELIDEA 777
 DB 718 MRLKETELRKVKVLGSAFGTVYKGIIPDGENVKIPVAIKVIRENTSPKANKELIDEA 777

DB 714 MRLKETELRKVKVLGSAFGTVYKGIIPDGENVKIPVAIKVIRENTSPKANKELIDEA 773
 QY 778 YVMAGVSSPVYSRLGLCLTSTVQLVYQVLTQMLPYCCLLDHRNENRGRLGSDLLMCMQIAK 837
 DB 774 YVMAGVSSPVYSRLGLCLTSTVQLVYQVLTQMLPYCCLLDHRNENRGRLGSDLLMCMQIAK 833
 QY 838 GMSYLEVDRVLRHDLAARNVLYSPNNKITTDFGLARLLDIDETEHYADGKVPKIMMAL 897
 DB 834 GMSYLEVDRVLRHDLAARNVLYSPNNKITTDFGLARLLDIDETEHYADGKVPKIMMAL 893
 QY 898 ESILRRFTTHQSDVMSYGVTVWELMTPGAKPYGIGIARIEIPDLLEKGERLPQPTCTIV 957
 DB 894 ESILRRFTTHQSDVMSYGVTVWELMTPGAKPYGIGIARIEIPDLLEKGERLPQPTCTIV 953
 QY 958 YMIWVKCMIDSECRPFRELVESEFRMARDPQRFVYIIONEDGPSAPDSTFYRSLLD 101
 DB 954 YMIWVKCMIDSECRPFRELVESEFRMARDPQRFVYIIONEDGPSAPDSTFYRSLLD 101
 QY 1018 DDMGDLVDAEYLVPOQGFPCPDPAQAGMVRHRRSSSTRSGGDLTLGLEPSEEAR 107
 DB 1014 DDMGDLVDAEYLVPOQGFPCPDPAQAGMVRHRRSSSTRSGGDLTLGLEPSEEAR 107
 QY 1078 RSPLAPSEAGSDVFDGDLGMGAAGLQSLPTHPDPSPLQRYSDPTVPLPSNDGVYAPL 113
 DB 1074 RSPLAPSEAGSDVFDGDLGMGAAGLQSLPTHPDPSPLQRYSDPTVPLPSNDGVYAPL 113
 QY 1138 TCSPOPEYVNOVDVPPSPREGPLPARPAGATLERAKTISPGKNGVYKQVAFAGAV 119
 DB 1134 ACSQPEYVNOVDVPPSPREGPLPARPAGATLERAKTISPGKNGVYKQVAFAGAV 119
 QY 1198 ENPEYLLPQGAAPQPPPAFSAFENLYYNOQDPPERKAPSTKGTPTANPEYGL 125
 DB 1194 ENPEYLLPQGAAPQPPPAFSAFENLYYNOQDPPERKAPSTKGTPTANPEYGL 125
 QY 1258 DVPV 1261
 DB 1254 DVPV 1257
 RESULT 3
 ERB2 MESAU STANDARD; PRT; 1254 AA.
 ID ERB2 MESAU AC 06053;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
 GN (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
 OS ERB2 OR NEU.
 OC Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Nerve;
 RX MEDLINE=94193007; PubMed=7908275;
 RA Nakamura T, Ushijima T, Ishizaka Y, Nagao M, Arai M,
 RA Yamazaki Y, Ishikawa T;
 RT "Cloning and activation of the Syrian hamster neu proto-oncogene."
 RL Gene 140:251-255(1994).
 CC - FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
 ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
 POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY BGF, TGF-
 ALPHA AND AMPHIREGULIN (BY SIMILARITY).
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 tyrosine phosphate.
 CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 (POTENTIAL).
 CC - SUBCELLULAR LOCATION: Type I membrane protein.
 CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE

CC RESIDUES.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D16285; BA03601.1; -.
 CC HSSP: P11362; IFCK.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_kinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR Pfam: PF02757; YLP; 2.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; FU; 3_pkinase; 1.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Proto-oncogene; Disease mutation.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 1254 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
 FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 653 675 POTENTIAL.
 FT DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 158 368 CYS-RICH.
 FT DOMAIN 472 644 CYS-RICH.
 FT DOMAIN 720 987 PROTEIN KINASE.
 FT NP_BIND 726 734 ATP (BY SIMILARITY).
 FT BINDING 753 753 ATP (BY SIMILARITY).
 FT ACT_SITE 845 845 BY SIMILARITY.
 FT DISULFID 195 204 BY SIMILARITY.
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 236 244 BY SIMILARITY.
 FT DISULFID 240 252 BY SIMILARITY.
 FT DISULFID 255 264 BY SIMILARITY.
 FT DISULFID 268 295 BY SIMILARITY.
 FT DISULFID 299 311 BY SIMILARITY.
 FT DISULFID 315 331 BY SIMILARITY.
 FT DISULFID 334 338 BY SIMILARITY.
 FT DISULFID 511 520 BY SIMILARITY.
 FT DISULFID 515 528 BY SIMILARITY.
 FT DISULFID 531 540 BY SIMILARITY.
 FT DISULFID 544 560 BY SIMILARITY.
 FT DISULFID 563 576 BY SIMILARITY.
 FT DISULFID 567 584 BY SIMILARITY.
 FT DISULFID 587 596 BY SIMILARITY.
 FT DISULFID 600 623 BY SIMILARITY.
 FT DISULFID 626 634 BY SIMILARITY.
 FT DISULFID 630 642 BY SIMILARITY.
 FT MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 530 530 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 571 571 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 658 658 V -> E (IN ONCOGENIC NEU).
 FT VARIANT 659 659 V -> E (IN ONCOGENIC NEU).
 SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2B51 CRG64;

Query Match 86.3%; Score 5916.5; DB 1; Length 1254;
 Best Local Similarity 86.3%; Pred. No. 7,3e-109;
 Matches 1089; Conservative 66; Mismatches 98; Indels 9; Gaps 3

QY 1 MELAALCRWGLLALLPPGAASVCTGTDMKRLPLASPEETHLMDLAHLYGQVQVGNL 60
 DB 1 MELAMCWMGLLALLSPGASGTCVCGTDMKRLPLASPEETHLMDLAHLYGQVQVGNL 60
 QY 61 ETTYPTNASLSPFODIOEVGYVLIANOVROVPLRLIVRGTOLEFEDNYALAVDNG 120
 DB 61 ETTYPTNASLSPFODIOEVGYVLIANOVROVPLRLIVRGTOLEFEDNYALAVDNR 120
 QY 121 DELNNTPTVGTASPGGLREQLRSLEILKGGVLIQRNPOLCYQDTILMKDIFHKNNQLA 180
 DB 121 DELNNTPTVGTASPGGLREQLRSLEILKGGVLIQRNPOLCYQDTILMKDIFHKNNQLA 180
 QY 121 DELDVTATATGTEPEGLREQLRSLEILKGGVLIQRNPOLCYQDTILMDVFRKNNQLA 180
 DB 121 DELDVTATATGTEPEGLREQLRSLEILKGGVLIQRNPOLCYQDTILMDVFRKNNQLA 180
 QY 181 LTLIDPTNSRACHPSPMCKSRGSSSEDCQSLTFTVACAGCARCKGLPLDDCHEQC 240
 DB 181 PVDIDTNSRACHPSPMCKSRGSSSEDCQSLTFTVAPRAVPAARLPLDDCHEQC 240
 QY 241 AAGCTGPRHSDCLACIHNHSGICELCPALVTYNTDTPESMPNDEGRYTFGASCVTACP 300
 DB 241 AAGCTGPRHSDCLACIHNHSGICELCPALVTYNTDTPESMPNDEGRYTFGASCVTACP 300
 QY 301 YNYLSTDVSGCTLYCPPLHNSVTLAEDGTQRCCKSKPCARVCGLGKFNFTVSPWLRLVP 360
 DB 301 YNYLSTDVSGCTLYCPPLHNSVTLAEDGTQRCCKSKPCARVCGLGKFNFTVSPWLRLVP 360
 QY 301 YNYLSTDVSGCTLYCPPLHNSVTLAEDGTQRCCKSKPCARVCGLGKFNFTVSPWLRLVP 360
 DB 301 YNYLSTDVSGCTLYCPPLHNSVTLAEDGTQRCCKSKPCARVCGLGKFNFTVSPWLRLVP 360
 QY 361 K-VSASHLEEFAGCKIPGSLAFIPESPDGPANNTPLQPEGLQVETLEITGYLYIS 410
 DB 361 K-VSASHLEEFAGCKIPGSLAFIPESPDGPANNTPLQPEGLQVETLEITGYLYIS 410
 QY 361 K-VSASHLEEFAGCKIPGSLAFIPESPDGPANNTPLQPEGLQVETLEITGYLYIS 410
 DB 361 K-VSASHLEEFAGCKIPGSLAFIPESPDGPANNTPLQPEGLQVETLEITGYLYIS 410
 QY 420 AMPDLSPLSVFONLQVIRGRILHNGAVSLTLOGISWGLRSRLSGSLAIHNNHT 470
 DB 420 AMPDLSPLSVFONLQVIRGRILHNGAVSLTLOGISWGLRSRLSGSLAIHNNHT 470
 QY 420 AMPDLSPLSVFONLQVIRGRILHNGAVSLTLOGISWGLRSRLSGSLAIHNNHT 470
 DB 420 AMPDLSPLSVFONLQVIRGRILHNGAVSLTLOGISWGLRSRLSGSLAIHNNHT 470
 QY 480 LCPVTVTWMDLFRPHALHTANREDEVGSLACHQLCARGHWGSGPQCVNCSQ 530
 DB 480 LCPVTVTWMDLFRPHALHTANREDEVGSLACHQLCARGHWGSGPQCVNCSQ 530
 QY 480 LCPVTVTWMDLFRPHALHTANREDEVGSLACHQLCARGHWGSGPQCVNCSQ 530
 DB 480 LCPVTVTWMDLFRPHALHTANREDEVGSLACHQLCARGHWGSGPQCVNCSQ 530
 QY 474 LCFVHTVWMDLFRPHALHTANREDEVGSLACHQLCARGHWGSGPQCVNCSQ 530
 DB 474 LCFVHTVWMDLFRPHALHTANREDEVGSLACHQLCARGHWGSGPQCVNCSQ 530
 QY 540 FLRGQCEVEGRVLOGLPREYNARHCLPCHPECCOPNGSVTCGFPADQVACAHKDP 590
 DB 540 FLRGQCEVEGRVLOGLPREYNARHCLPCHPECCOPNGSVTCGFPADQVACAHKDP 590
 QY 534 FLRGQCEVEGRVLOGLPREYNARHCLPCHPECCOPNGSVTCGFPADQVACAHKDP 590
 DB 534 FLRGQCEVEGRVLOGLPREYNARHCLPCHPECCOPNGSVTCGFPADQVACAHKDP 590
 QY 594 PFCVAPRCPSGVKPDLSYPIKYPDEBGMCPDINCTHSCVDLDERGCPAEGRASPTS 650
 DB 594 PFCVAPRCPSGVKPDLSYPIKYPDEBGMCPDINCTHSCVDLDERGCPAEGRASPTS 650
 QY 600 PFCVAPRCPSGVKPDLSYPIKYPDEBGMCPDINCTHSCVDLDERGCPAEGRASPTS 650
 DB 600 PFCVAPRCPSGVKPDLSYPIKYPDEBGMCPDINCTHSCVDLDERGCPAEGRASPTS 650
 QY 654 IYSAVVGILVAVGVGILIKRROOKIRKYTRRLLOETELVEPLTPSGAMPNOQMR 710
 DB 654 IYSAVVGILVAVGVGILIKRROOKIRKYTRRLLOETELVEPLTPSGAMPNOQMR 710
 QY 720 ILKETELRKVKVLSGAGFYKGIWIPDGENVKIPIAIVLRENTSPKANKEILDEAYV 770
 DB 720 ILKETELRKVKVLSGAGFYKGIWIPDGENVKIPIAIVLRENTSPKANKEILDEAYV 770
 QY 714 ILKETELRKVKVLSGAGFYKGIWIPDGENVKIPIAIVLRENTSPKANKEILDEAYV 770
 DB 714 ILKETELRKVKVLSGAGFYKGIWIPDGENVKIPIAIVLRENTSPKANKEILDEAYV 770
 QY 780 MAGVSPFVSLGLCLTSTVQVLTQMLPYGCLLDHREHGRGSDLLNCVQIAKGM 830
 DB 780 MAGVSPFVSLGLCLTSTVQVLTQMLPYGCLLDHREHGRGSDLLNCVQIAKGM 830
 QY 774 MAGVSPFVSLGLCLTSTVQVLTQMLPYGCLLDHREHGRGSDLLNCVQIAKGM 830
 DB 774 MAGVSPFVSLGLCLTSTVQVLTQMLPYGCLLDHREHGRGSDLLNCVQIAKGM 830
 QY 840 SYLEAVRLVHDLAARNVLVSPNHVKTITDGLARLLDIDETEYHAAGKVPIMKMALES 890
 DB 840 SYLEAVRLVHDLAARNVLVSPNHVKTITDGLARLLDIDETEYHAAGKVPIMKMALES 890
 QY 834 SYLEAVRLVHDLAARNVLVSPNHVKTITDGLARLLDIDETEYHAAGKVPIMKMALES 890
 DB 834 SYLEAVRLVHDLAARNVLVSPNHVKTITDGLARLLDIDETEYHAAGKVPIMKMALES 890
 QY 900 ILRRFTTHQSDWVSQVTVWELMTFGAKYDGLIARBEPLDLEGESELPPOPICITDIYVM 950
 DB 900 ILRRFTTHQSDWVSQVTVWELMTFGAKYDGLIARBEPLDLEGESELPPOPICITDIYVM 950
 QY 894 ILRRFTTHQSDWVSQVTVWELMTFGAKYDGLIARBEPLDLEGESELPPOPICITDIYVM 950
 DB 894 ILRRFTTHQSDWVSQVTVWELMTFGAKYDGLIARBEPLDLEGESELPPOPICITDIYVM 950
 QY 960 IIVKCMWIDSECRPRFELVSEFARMARDPQRFVIVIQNEIDLGSPSPIDSTFFYSLEDDD 1010
 DB 960 IIVKCMWIDSECRPRFELVSEFARMARDPQRFVIVIQNEIDLGSPSPIDSTFFYSLEDDD 1010
 QY 954 IIVKCMWIDSECRPRFELVSEFARMARDPQRFVIVIQNEIDLGSPSPIDSTFFYSLEDDD 1010
 DB 954 IIVKCMWIDSECRPRFELVSEFARMARDPQRFVIVIQNEIDLGSPSPIDSTFFYSLEDDD 1010

QY 1020 MGDLYDAEYLVPOQGFPCPDAPAGAGMHHRRSSSTRSGGDLTLGLEPSEEEAPRS 1079
 |||||
 DB 1014 MGDLYDAEYLVPOQGFPCPDAPAGAGMHHRRSSSTRSGGDLTLGLEPSEEEAPRS 1073
 |||||
 QY 1080 PLAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSFLQYSSDPFVPLPSETDGYVAPLTC 1139
 |||||
 DB 1074 PLAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSFLQYSSDPFVPLPSETDGYVAPLTC 1133
 |||||
 QY 1140 SPQPEYVNOQDVPRPQPPREPEPLPAAPAPAGATLEPRATLSFGKNGYKDVFAFGAVEN 1199
 |||||
 DB 1134 SPQPEYVNOQDVPRPQPPREPEPLPAAPAPAGATLEPRATLSFGKNGYKDVFAFGAVEN 1193
 |||||
 QY 1200 PEYLTPQGGAPQHPHPPASPAFNLVYWDQDPERGAPESTKGTPTAENPEYGLDV 1259
 |||||
 DB 1194 PEYLTPQGGASQPH-PPALCPAFNLVYWDQDPERGAPESTKGTPTAENPEYGLDV 1252
 |||||
 QY 1260 PV 1261
 |||||
 DB 1253 PV 1254
 |||||
 RESULT 4
 EGFR_HUMAN STANDARD: PRT: 1210 AA.
 ID EGFR_HUMAN
 AC P06533; P06268; Q14225; Q9JMD7; Q8UMD8; Q9JMG5; Q92795; Q00732;
 AC 000688; Q9B2S2; Q9H2C9; Q9GZK1; Q9H3C9;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
 DE protein-tyrosine kinase ErbB-1).
 GN EGFR OR ERBB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_Taxid:9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=84219729; PubMed=6328312;
 RA Ulrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
 RA Lee J., Yarden Y., Liberman T.A., Schlessinger J., Downward J.,
 RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant
 RT expression of the amplified gene in A431 epidermoid carcinoma cells";
 RL Nature 309:418-425(1984).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RA Ilekis J.V., Stark B.C., Scoccia B.;
 RT "Possible role of variant RNA transcripts in the regulation of
 RT epidermal growth factor receptor expression in human placenta";
 RL Mol. Reprod. Dev. 41:149-156(1995).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RA MEDLINE=97078686; PubMed=8918811;
 RA Reiter J.L., Maible N.J.;
 RT "A 1.8 kb alternative transcript from the human epidermal growth
 RT factor receptor gene encodes a truncated form of the receptor";
 RL Nucleic Acids Res. 24:4050-4056(1996).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RA MEDLINE=97256547; PubMed=9103388;
 RA Ilekis J.V., Garliti J., Niederberger C., Scoccia B.;
 RT "Expression of a truncated epidermal growth factor receptor-like
 RT protein (TBEGR) in ovarian cancer";
 RL Gynecol. Oncol. 65:36-41(1997).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
 RX TISSUE=Placenta;
 RA MEDLINE=21100872; PubMed=1161793;
 RX

RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
 RA Schehl Sinclair C., Pearshall R.S., Green P.J., Yee D., Lampland A.L.,
 RA Balasubramanian S., Crossley T.D., Magnuson T.R., James C.D.,
 RA Maible N.J.;
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative EGFR transcripts encoding truncated receptor
 RT isoforms";
 RL Genomics 71:1-20(2001).
 RN [6]
 RP SEQUENCE OF 575-687 FROM N.A.
 RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,
 RA Lampland A.L., Balasubramanian S., Crossley T.O., Magnuson T.R.,
 RA Maible N.J.;
 RT "Human and mouse alternative EGFR transcripts encoding only the
 RT extracellular domain of the receptor";
 RL Submitted (Feb-1999) to the EMBL/Genbank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 713-924 FROM N.A.
 RX MEDLINE=84196372; PubMed=6326261;
 RA Lin C.R., Chen W.S., Krueger W., Stolarek L.S., Weber W.,
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
 RT "Expression cloning of human EGF receptor complementary DNA: gene
 RT amplification and three related messenger RNA products in A431
 RT cells";
 RL Science 224:843-848(1984).
 RN [8]
 RP SEQUENCE OF 150-962 FROM N.A.
 RX MEDLINE=84424583; PubMed=6330563;
 RA Xu Y.H., Ishii S., Clark A.U.L., Sullivan M., Wilson R.K., Ma D.P.,
 RA Roe B.A., Merlino G.T., Pastan I.;
 RT "Human epidermal growth factor receptor cDNA is homologous to a
 RT variety of RNAs overproduced in A431 carcinoma cells";
 RL Nature 309:806-810(1984).
 RN [9]
 RP SEQUENCE OF 1028-1210 FROM N.A.
 RX MEDLINE=85046483; PubMed=6093780;
 RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
 RA O'Malley B.W.;
 RT "Isolation of an evolutionarily conserved epidermal growth factor
 RT receptor cDNA from human A431 carcinoma cells";
 RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
 RN [10]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=88217333; PubMed=3329716;
 RA Haley U.D., Whittle N., Bennett P., Kinchington D., Ulrich A.,
 RA Waterfield M.D.;
 RT "The human EGF receptor gene: structure of the 110 kb locus and
 RT identification of sequences regulating its transcription";
 RL Oncogene Res. 1:375-396(1987).
 RN [11]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=91107677; PubMed=1988448;
 RA Haley U.D., Waterfield M.D.;
 RT "Contributory effects of de novo transcription and premature
 RT transcript termination in the regulation of human epidermal growth
 RT factor receptor proto-oncogene RNA synthesis";
 RL J. Biol. Chem. 266:1746-1753(1991).
 RN [12]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=85270438; PubMed=2291699;
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
 RT "Characterization and sequence of the promoter region of the human
 RT epidermal growth factor receptor gene";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
 RN [13]
 RP SEQUENCE OF 540.
 RA Konda D.;
 RL Submitted (SEP-1997) to the SWISS-PROT data bank.
 RN [14]
 RP RECEPTOR ACTIVITY.
 RX MEDLINE=84191554; PubMed=6325248;
 RA Mroczkowski B., Mosig G., Cohen S.;
 RT "ATP-stimulated interaction between epidermal growth factor receptor

RT and supercoiled DNA.";
 RL Nature 309:270-273(1984).
 RN [15]
 RP PHOSPHORYLATION.
 RX MEDLINE=89278137; PubMed=2543678;
 RA Maxolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.W.,
 RA Howk R., Givol D., Ullrich A., Schlessinger J.;
 RT "All autophosphorylation sites of epidermal growth factor (EGF)
 RT receptor and HER2/neu are located in their carboxyl-terminal tails.
 RT Identification of a novel site in EGF receptor.";
 RL J. Biol. Chem. 264:10667-10671(1989).
 RN [16]
 RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
 RP ASN-528.
 RX MEDLINE=96398132; PubMed=8962717;
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
 RT "Analysis of the glycosylation patterns of the extracellular domain of
 RT the epidermal growth factor receptor expressed in Chinese hamster
 RT ovary fibroblasts.";
 RL Growth Factors 13:121-132(1996).
 RN [17]
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
 RP ASN-603.
 RX MEDLINE=20198209; PubMed=10731668;
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
 RT "Characterization of the N-glycosaccharides attached to the atypical
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor
 RT receptor.";
 RL J. Biochem. 127:65-72(2000).
 RN [18]
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE=98225196; PubMed=9556602;
 RA Abe Y., Oda M., Inagaki F., Lax I., Schlessinger J., Kohda D.;
 RT "Disulfide bond structure of human epidermal growth factor.";
 RL J. Biol. Chem. 273:11150-11157(1998).
 RN [19]
 RP REVIEW.
 RX MEDLINE=87297456; PubMed=3039909;
 RA Carpenter G.;
 RT "Receptors for epidermal growth factor and other polypeptide
 RT mitogens.";
 RL Annu. Rev. Biochem. 56:881-914(1987).
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
 CC EGF-like growth factor, Gp30 and vaccinia virus growth factor. Is
 CC involved in the control of cell growth and differentiation.
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
 CC secreted.
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/
 CC truncated isoform/TSGFR, 3/p110 and 4; are produced by
 CC alternative splicing.
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
 CC expressed in ovarian cancers.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X00588; CAA25240.1; -;
 DR EMBL; U95089; AAB51063.1; -;
 DR EMBL; U48722; AAC50802.1; -;

DR EMBL; U48723; AAC50804.1; -;
 DR EMBL; U48724; AAC50796.1; -;
 DR EMBL; U48725; AAC50797.1; -;
 DR EMBL; U48726; AAC50798.1; -;

Query Match 45.9%; Score 3148; DB 1; Length 1210;
 Best Local Similarity 49.6%; Pred. No. 6e-161;
 Matches 631; Conservative 177; Mismatches 351; Indels 112; Gaps 2;

11 LLLALPPGAA--STOVCTGDMLELRLPASPEHLDMLRHLVGGCCVGVGNLTETLPTN 68
 14 LLAALCPRLALEKRYKQVQSTSNKLTQGTEDFELSQQFNNCEVVLGNLEITTYQRN 73
 69 ASLSFLDIDQEVGYVLIANNOVQVPLQRLIRIVRGTQLEFDNYALAVLNDGPPLNNTTP 124
 74 YDLSFLTKIQEVAGVLIALTVERIPLENQIRGNMYENSVALAVLSYND----- 124
 129 VTGASPGRLRLQLRSLTEILKGVLIQRPOLCYOTILMKDIFKGNOLALTLIDTR 181
 127 --ANKTELKELPMKNLQETILGAVRESNNPALCNVSIQMRDIVSSDFLSNMSMDQNH 181
 189 SRACPCSPMKGSRGSGESSEDCSLTRTVACAGCA-RCKGPLPTDCHGQCAAGCTGP 241
 184 LGSQCKDPCSPNGSCMGAGSENGKLTIIICAQCCSGRCGRKSPDCCNQCAGCTGP 241
 248 KHSQCLALAHNHSIGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACPYNYLSTD 301
 244 RESDLYCRKFRDEATCKDCTPPMLNPTTYQMDVNPPEGYSFGATCVKCKCPANVYTD 301
 308 VGSCTLVCLPLNNOETVADGTORCEKSKPCARCYGLGFNNFTVFMLRVVPVASHL 361
 304 HSCVTRACGADSYEN-EDGYRKKCKGEGCRKRVNGIGI-GEFYKSL-----SINKATNI 356
 368 EFAPACKTIFGSLAFLPESPDPDASNTAPLQBPQLQVFTELEITGYLYISAWPSLDP 421
 357 KHFKCTSIQGDHLPLVAFRGSDFTHTPPLDPELIDLTKEITGFLLIQAMPERTD 414
 428 LSVFONLOVITGRILLNAGAVSLTQGLSIGSLGRLSRLSGSLALHHNTILCFVHTVP 481
 417 LHAFFNLIELIGRKTQHQGFSLAVASLNTISLGRSLKESIDGVIITISGKNLCYANTIN 474
 488 WDLFRNHOALLHTANRPEDECEVGEGLACQHLQCARHGCGPGTQCVCNSQPLRQGEV 541
 477 WKGLFSGTGQTKTIISNRGENSCATQOVCHALCSPEGCMGPEPRDVCNRNVRGECV 534
 548 EECRYLQCLPEEYVNAHRCPCHEPCOPQNSVTCSFPEADQCVAAHYDPPFCVACRP 601
 537 DKCNLLEGEPRFENSECIQCHEPCLPQAMNITCTGRGPDNCTQCAHYIDGPHCVKTCP 596
 608 SGVRKDSLYMELWKFPEDEGACQPCPLNCTHSCVDLDDKCGPAGRASPLTSIVSAVVG- 664
 597 AGWGEENNTL-VMKYADGHVCHLCHPCTCTGCTGPGLEGCTPNGKPIP--STATGVGA 651
 667 --ILLVVVLGVVFGILIKRQOKIRKXTMRLLQETELVLEPLTPSGAMPQAQRILKET 724
 654 LLLLVALLVG---LFMRRHIVRKTLLRLDERELVEPLTPSGAPNALRLIKET 710
 725 ELRKKTUGSAPFGVYKGIITPGENVVKIPVAKVLRNTSPRANKETIDEAIVYVAGVG 784
 711 EFKIKVIGSAPFGVYKGIIPGEKVKIPVALKELRATSPRANKETIDEAIVYVAVSD 770
 785 SPVYSRLIGLICLTSTVQLVTLQMLPYGCLLHVRENRRLSSODLLNMCQIAGMSYLED 844
 771 NPHVRRLIGLICLTSTVQLVTLQMLPYGCLLHVRENRRLSSODLLNMCQIAGMSYLED 830
 845 VLVHRLDLAANNVLKSBHVKITDPGLARLIDETEPHADGKVPDKMALESILRRR 904
 831 RLVHRLDLAANNVLKSBHVKITDPGLARLIDETEPHADGKVPDKMALESILHRI 890
 905 FTHGSDVNSYGVATWELMTFGAKPYDGIIPAREPDLLEKGERLPPOPPICTIDVYMIWVK 964
 891 YTHGSDVNSYGVATWELMTFGSKPYDGIIPASEISITUEKERLPPOPPICTIDVYMIWVK 950


```

FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 510 523 BY SIMILARITY.
FT DISULFID 526 535 BY SIMILARITY.
FT DISULFID 539 555 BY SIMILARITY.
FT DISULFID 558 571 BY SIMILARITY.
FT DISULFID 562 579 BY SIMILARITY.
FT DISULFID 562 591 BY SIMILARITY.
FT DISULFID 562 591 BY SIMILARITY.
FT DISULFID 595 617 BY SIMILARITY.
FT DISULFID 620 628 BY SIMILARITY.
FT DISULFID 624 636 BY SIMILARITY.
FT MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1112 1112 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE) (BY SIMILARITY).
FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 4).
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46DF2D2F5 CRC64;

```

Query Match 45.6%; Score 3124; DB 1; Length 1210;
 Best Local Similarity 49.6%; Pred. No. 1,le-159;
 Matches 635; Conservative 172; Mismatches 353; Indels 120; Gaps 26;

```

QY 11 LLLALLPPGAA--STOVCTGTDMLKRLPASBETHLDMRLHYQCGVYQGNLELYLPTN 68
DB 14 LTLTLCAGAGALEEKVCQGTNSRLTQGTEDHFLSLQSNYNNCVELVGNLEITLYQGR 73
QY 69 ASLSFLQDIQEVQGVYLIANQVRQVPLRLIRYRGQLFEDNYALAVLDNGPLNNTPP 128
DB 74 YDLSTLKTIOEVAGVYLIANTVERIPLENIQIRGALYENTYALALISN----- 124
QY 129 VTGASPGGLREIOLRSTLEIKGVLIOBNPOLCYODTILMKDI---FHKNQALTLTL 184
DB 125 -YGTNRGTGRELPKRNLOEILIGAVRFSNNPILCNMTIQRDLYQVVEFNNMSMDL--- 180
QY 185 DTNRSRACHPSCPMCKGSRGWESSEDCQSLTRTVACAGCA-RCKGPLPTDCCHEDQCAAG 243
DB 181 -QSHPSSCPCKDCPSPNCSGWGGEENCQKLTIRICAQCCSHRCRGRSPSDCCNQCAG 239
QY 244 CTGPKHSDCLACHFNHSGICELHCPALVYNTTFESMENPBERITYFGASCYTACPRNY 303
DB 240 CTGPRESDCLVQCFODEATCKDTPMLNPTTYMDVNPBGKVSFGATCVKCKPRNY 299
QY 304 LSTDVSGCTLVCPHNOEVTAEQTORCEKCSKPCARVCYGLGW--FNNFTVSPMLRVBK 361
DB 300 VYTDHSGCVRAQCPGYEV--BEDGIRCKCKDDGCRVAVNGIGIGERKD-TLS----- 350
QY 362 VSASHLEEFACCKIFGSLALPSSFGDDPSANTAPLQPOLOVFEETLEITLYLYTAM 421
DB 351 INATNIKHFYCTAISGDLHLTPAFKGDSEFTTRPPDPRLEILTKVXKITITFLLIQAM 410
QY 422 PDSLPLDSVFNLOVIRGRILHNGAYSLTQGLISWLGRLRELGSGALILHNTHLC 481
DB 411 PDNMTDLHAFENLEIRGRITKHQHFSLAVGINTISLGRLSKETISDGVITISGRNLC 470
QY 482 FVHTVPMQDLFRNPHQALLHTANFEDCECVBGLACHQLCARHGCWGPGEFTQCVCNSQFL 541
DB 471 YANTINMKKLLFGTNPQKTKIMNRAEKCAVNVHVCNPLCSSGSGCPEPRDCVSCONS 530
QY 542 RGQECVEECVILQGLPREYVVARHCLPCHCECQPNQNSVTCFPEBADQCVACAHYNDPFF 601

```

```

DB 531 RGRECEVCENILLEGPREFVENSECIQCHEBCLPQANNTCTGGRPONCIQCAHYIDGPH 590
QY 602 CVARPSGVKDDLSWPIWKFPDEBGAQCPPICTNCHSCYDLDDKGRPARQASPLSTIV 66;
DB 591 CVKTECPAGIMKENNTL-VWKYADANNVCHLCHACTGACGPGIQCCEVWPSGKIPISA 64;
QY 662 SAVVGLLVVLGVVFGI-LIKRQOKIKRYKWRRLQETELVEPLTPSGAMPQAOMRI 720
DB 650 TGIIVGLFLIV-VALGILFMRNRHIVRKTILARLLQERELVEPLTPSGAPRQALRI 700
QY 721 LKETELRKYVLGSGARCTYKGIWIDGENVKIPVAIKLRENTSPKANKELIDEXYVM 780
DB 709 LKETEFKKIKVLGSGAFGTIVKGIWIDGENVKIPVAIKELRENTSPKANKELIDEXYVM 760
QY 761 AGVSPYVSRLLGCLTSTVOLVQLMPYGLDHDVENVGRGLSDOOLMMCMQIAGMS 840
DB 769 ASVDNPHCRLLGICLSTVOLITQMLPFGCLDYREHKNDISOYLMMCVQIAGMN 820
QY 841 YLEDVRLVHRDLAARNVLYKSPNFKITDQGLARLIDIDETVHADGKVPKIMMALESI 900
DB 829 YLEDVRLVHRDLAARNVLYKTPQVKITDQGLARLIDIDETVHADGKVPKIMMALESI 880
QY 901 LRRRTQSDVWSYGVYVWELMTGAPRYDIPAREIPDLLEKGERLPQPPICITIDYMI 960
DB 889 LRRITQSDVWSYGVYVWELMTGSKPYDIPASDISILEKGERLPQPPICITIDYMI 940
QY 961 MWKCMIMIDSECRPRELIVSEFSSMARDPORFVYIQ-NEDLGPRASPLDSFFYSILEDDO 101
DB 949 MWKCMIMIDADRRPFRRELILEFSQADDPQRYLVITQDDEMHLPSPIDSNFYALMDEED 100
QY 1020 MGDLVDAEYVLPQCGFCPPDPAFAGAGVYHHRSSSTRSGGDLTLGLEPSEEBAPRS 107
DB 1009 MEDVVDADYELIPQCGF-----NSPST-----SRT 103
QY 1080 PLAPSEAGSDVPFGDGLGMAKAGLQSLPHDPSRLORYSEDPTRPLPSET--DGYVAPL 113
DB 1035 PLSSLSATSN---NSTVACINRNSCRKEDAFRLRYSSDPDGAVENTIDAFU--- 106
QY 1138 TCSPOPEYVNPDPVPPSPBREGPLPAPAPAGATLERAKTISPGRKGVYVYKDFAFAGAV 115
DB 1088 --PVPEYVNG-SVPKBPAGSVQPVYVHNPRLP-----APGRDLHYQN--PSNAY 113
QY 1138 ENPEYL-TPQGAAPQCHPPAPSPADNLYYDQ-----DP-----PCKGAP 124
DB 1134 GNPEYLTNAQ-----PTCLSSGGSFSLWLOKSHQMSLDNPDYQDFFPKETKPN 118
QY 1241 STFKQTPAENPEYLGDPV 1260
DB 1185 GIFKG-PTAENAEYLRYVAP 1203

```

RESULT 6
 ERB4_HUMAN STANDARD; PRT; 1308 AA.
 ID ERB4_HUMAN
 AC Q15303;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-4 precursor (BC 2.7.1.112)
 DE (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
 GN ERB4 OR HER4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM JM-A).
 RC TISSUE=Breast carcinoma;
 RX MEDLINE=91189574; PubMed=8383326;
 RA Ploeman G.D., Choussou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
 Foy L., Neubauer M.G., Shoyab M.,
 "Ligand-specific activation of HER4/p180erbB4, a fourth member of the

RT epidermal growth factor receptor family.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
 RC TISSUE=Fetal brain;
 RX MEDLINE=97476287; PubMed=9334263;
 RA Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plozman G.D.,
 RA Klagsbrun M.;
 RT "A novel juxtamembrane domain isoform of HER4/ErbB4. Isoform-specific
 tissue distribution and differential processing in response to
 RT Phorbol ester.";
 RL J. Biol. Chem. 272:26761-26768(1997).
 CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
 CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
 CC NTA. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
 CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
 CC RECEPTORS (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; JM-A (SHOWN HERE) AND JM-B;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
 CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER. JM-A IS PROCESSED
 CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
 CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
 CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
 CC KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,
 CC PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,
 CC LUNG, SALIVARY GLAND, AND PANCREAS.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L07866; AAB59446.1; -.
 DR HSSP: P13622; JFCK.
 DR GeneW: HGNC:3432; ERBB4.
 DR MIM: 600543; -.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_Pkinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00069; Pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR Pfam: PF02757; YLP; 2.
 DR Prodom: PD000001; Euk_Pkinase; 1.
 DR SMART: SM00261; FU; 4.
 DR SMART: SM00219; TYRC; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane: Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Alternative splicing;
 FT SIGNAL 1 25
 FT CHAIN 26 1308 POTENTIAL.
 FT DOMAIN 26 651 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
 FT TRANSMEM 652 675 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 676 1308 POTENTIAL.
 FT DOMAIN 186 334 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 496 633 CYS-RICH.
 FT DOMAIN 718 985 PROTEIN KINASE.

FT NP BIND 724 732 ATP (BY SIMILARITY).
 FT BINDING 751 751 ATP (BY SIMILARITY).
 FT ACT SITE 843 843
 FT DISULFID 189 197 BY SIMILARITY.
 FT DISULFID 193 205 BY SIMILARITY.
 FT DISULFID 213 221 BY SIMILARITY.
 FT DISULFID 217 229 BY SIMILARITY.
 FT DISULFID 230 238 BY SIMILARITY.
 FT DISULFID 234 246 BY SIMILARITY.
 FT DISULFID 249 258 BY SIMILARITY.
 FT DISULFID 262 289 BY SIMILARITY.
 FT DISULFID 293 304 BY SIMILARITY.
 FT DISULFID 308 323 BY SIMILARITY.
 FT DISULFID 326 330 BY SIMILARITY.
 FT DISULFID 503 512 BY SIMILARITY.
 FT DISULFID 507 520 BY SIMILARITY.
 FT DISULFID 523 532 BY SIMILARITY.
 FT DISULFID 536 552 BY SIMILARITY.
 FT DISULFID 555 569 BY SIMILARITY.
 FT DISULFID 559 577 BY SIMILARITY.
 FT DISULFID 580 589 BY SIMILARITY.
 FT DISULFID 593 614 BY SIMILARITY.
 FT DISULFID 617 625 BY SIMILARITY.
 FT DISULFID 621 633 BY SIMILARITY.
 FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 138 138 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 174 174 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 181 181 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 253 253 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 358 358 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 473 473 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 548 548 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 620 620 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARSPLIC 626 648 NGPSTHCITYPMWGHSLPQHA -> IGSSIEDCIGLMD
 FT (IN ISOFORM JM-B).
 SQ SEQUENCE 1308 AA; 146807 MW; 5E4AE80968D88761 CRC64;
 Query Match 43.5%; Score 2981.5; DR 1; Length 1308;
 Best Local Similarity 45.2%; Pred. No. 5.1e-152;
 Matches 611; Conservative 187; Mismatches 375; Indels 179; Gaps 29
 QY 9 WGLIALLPQAA-----STGYCTGTDKMLRLPASPEETHLMDLRHLYGCGVQGNLELTY 64
 DB WWWSLVVAAGTVPSPDSQSCVCACTENKSLSLDLEQYBALRKYYENGCVGNLEITS 67
 QY 65 LPTNASLSFLDDIOEVQGYVLIANOVROYPLQRLRIVRGTOLFEDNYALAVLDNGDPLN 124
 DB IEHNRDLSEFLRSVREYGVYVALNORRYPLELRLIRIRGTLYEBRYALAIPLNRRKOG 127
 QY 125 NTPVPTGASPGALRELOLRSLTEILKGVYLIORPOLCYODTILMKDIFKNNQALTLTI 184
 DB NF-----GLDELGKNTLTLNGSYVVDQKFLCYADTITHMDIVNPPNSULTLV 178
 QY 185 DTRSRACHPSGPKCKSGRCWGESSEDCQGLTRTVCAAGS-ARCKGPLPTDCHQECAG 243
 DB STNGSSGCGRCHKSCGTG-RCWGPTENHCQTLTRTVCAECCDGCYGGYVSDCCCHRCACAG 237
 QY 244 CTGPRHSGCLCLHPNHSIGIELHCPALVLYNPTDPSMNPBGRYTFGASCVTACPVNY 303
 DB CSGPDTICFCAMFNDSGALCVTCCPQTFYNNPTTFQLENFAKTIYGAFCYKCKPHNF 297
 QY 304 LSTDVGSCTLVCPILNDEVTAAEDGTQRCCKSKRCARVCYGLGMFNFTVSPFLRVPKYS 363
 DB V-VDSSSCVRAKCPASSKREV-EENGIKKCKECTDIPCACDGIG-----TGSLSAQTV 349
 QY 364 ASHLEEFQCKKIKGSLAFLPESDGDPAANTAPLQDEQVQVETLEITGYVYIAMPD 423

```

Db 350 SSNIDKFINCKINGNIIFLVGTIGHDPYNAIEAIDEKLNVTFRVTEITGILNIGSWMP 409
Qy 424 SLPLSVFQNLQVIRGRILHNGAYSLTLQGLGISMGLASRLREISGLALIHNTHLCEY 483
Db 410 NMTDFSVNSNLVLTIGRVLVYSGLSLLILKOGITSLQPSGLKEISAGNIITYTNSLCTY 469
Qy 484 HTVPMDOLEFNPHQALLHTANRPDECEVGEGLACHQUCARGHCWCGPGPTQCVNCSQFLRG 543
Db 470 HTIMWTLFFSTINORIVIRDNKKAENCTAEGMVCNHLSSDGCGPDPDCLSCRFRRSG 529
Qy 544 QCVBECEVGLQGLREYVNAHNCIPCHPEOP-ONGSVTGEADQCVACVAXKDPKPC 602
Db 530 RICHSCNLYGERRERFENGSIQECQDQCKMEDGLTLHGEPDNYKCSHFKQGPNC 589
Qy 603 VAPCSGVKPDLSYMPWKPFDEGACQCPINCTHSCVDLDDKGC-----PAE 651
Db 590 VEKCPDGLQGANSF--IFKVADPDRCHPCHPNCTQCGNPTSHDCLYPMWHTSLPHQ 647
Qy 652 QASPLTIVAVV-GLLVVVLGVVFGILKRCQKIRKYMRLLOETELVEP:TPSG 710
Db 648 AR-TPL--IAAGVIGGLFVLIVGLTAVVYRRKSIR-KKRALRFL-ETELVEP:TPSG 702
Qy 711 AMPNQAQRIKETEELKRVKVLGSGAFGVYKGIWIPDGENVXIPVAKVLRNTPSKAN 770
Db 703 TAPNQAQRIKETEELKRVKVLGSGAFGVYKGIWPEGETVAKIPVAKITLNTTGPKAN 762
Qy 771 KEILDEAVVAGVSGPYVSRLLGICLTSTVQVLTQMPYQCLLDHVENRGRIGSQDLIN 830
Db 763 VEFDEALIMASMHPLVRLGLCLSPITQLVQMPHQCILEYVEHEDNIGSOLLN 822
Qy 831 WMOJAKGNSYLEDRVLRDLAARNLVKSPNNVKTIDGGLRLDIDETEHAGRGV 890
Db 823 WVOJAKGNSYLEDRVLRDLAARNLVKSPNNVKTIDGGLRLDIDETEHAGRGV 882
Qy 891 PIKMMALLESILRRFTHQSDVWSYGVTVWELMTFGANPYDGIAPAREIPDLLEKGERLPOR 950
Db 883 PIKMMALLESILRRFTHQSDVWSYGVTVWELMTFGANPYDGIAPAREIPDLLEKGERLPOR 942
Qy 951 PICTIDVYMWKCMMDISECRFRFRVSEFSMADPORFVYVIONED-LGASLSDST 1009
Db 943 PICTIDVYMWKCMMDISECRFRFRVSEFSMADPORFVYVIONED-LGASLSDST 1002
Qy 1010 FYRSLIEDDDMDGLVDAEYLVPOQGFCCPDPAAGAGMWHHRSSSTPSGGDLTLGL 1069
Db 1003 FFQNLDEEDLEDMDADEVLPV-QAFNIPPP-----YTSARLSDSNS-----EIGH 1050
Qy 1070 EPSEEBAPRS-----PLAP-SEGAQSCVFPDGLMGKA 1100
Db 1051 SPAPAYTPMSGNQFVYRDGGAAGVSVYRAFTSTIPAPVAQGATAIFPDSCCNGT 1110
Qy 1101 AKGLQSLPTHDPSPLORYSEDPYVLPSS-----ETDGVVAPLITCSPOEYVNGQDVARP 1153
Db 1111 LRKVAPHVQEDSSTQKYSADPTVFAPEBSPRGSLDEGIMTMRDKPKKEYINPVE--- 1167
Qy 1154 QPSPRREGPLPAADPAKATLERAKTSLPGKGVVKKVDFAFGAVENPEYLTPOGGAAPOR 1213
Db 1168 -----ENPFVSRR-----KNGDLQ-----ALDNPVYHNASNG----- 1194
Qy 1214 HPPPA-----SPAFLNLYWDDQPPPERGA- 1238
Db 1195 -PFKADEYVNEPIYLNTPANTLCKAAYLKONNISMPKAKKAFDNDPNWNSLPRSTL 1253
Qy 1239 -PSTFKGTPT-----AENPEYL 1255
Db 1254 QHPDYLGQYSTKYFYKONGRIPIVAENPEYL 1265

```

```

RESULT 7
ERR4_RAT
ID ERR4_RAT STANDARD; PRT; 1308 AA.
AC Q62956; Q922N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)

```

```

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN ERB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
RA Marchionni W.A., Kelly R.A.;
RT "Neuregulins promote survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes."
RL J. Biol. Chem. 273:10261-10269 (1998).
RN (2)
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=9122560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system."
RL Neuron 6:691-704 (1991).
RN (3)
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Froehner P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration."
RL J. Neurosci. 17:1642-1659 (1997).
RN (4)
RP FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
RN 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
RN NTRK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
RN NOT ACTIVATED BY EGF, TGF- $\alpha$  AND AMPHIREGULIN (BY SIMILARITY).
RN CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
RN tyrosine phosphate.
RN SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
RN RECEPTORS (POTENTIAL).
RN SUBCELLULAR LOCATION: Type I membrane protein.
RN TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
RN NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
RN OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
RN RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
RN HEART.
RN -!- LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
RN RESIDUES (BY SIMILARITY).
RN -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
RN This SWISS-PROT entry is copyright. It is produced through a collaboratio
RN between the Swiss Institute of Bioinformatics and the EMBL outstation
RN at the European Bioinformatics Institute. There are no restrictions on it
RN use by non-profit institutions as long as its content is in no wa
RN modified and this statement is not removed. Usage by and for commercia
RN entities requires a license agreement (See http://www.isb-sib.ch/announce
RN or send an email to license@isb-sib.ch).
RN EMBL; AF041838; AAC08899.1; -.
RN EMBL; U52531; AAC53051.1; -.
RN HSSP; P11362; IFKG.
RN InterPro: IPR0006494; EGFR_L domain.
RN InterPro: IPR0007219; Euk_Dkinase.
RN InterPro: IPR002174; Furin-like.
RN InterPro: IPR001245; Tyr_Dkinase.
RN InterPro: IPR004019; YLP_motif.
RN Pfam; PF00757; Furin-like; 1.
RN Pfam; PF00069; Pkinase; 1.
RN Pfam; PF01030; Recep_L_domain; 2.
RN Pfam; PF02757; YLP_2.
RN PRINTS; PR00109; TYRKINASE.
RN ProDom; PD000001; Euk_pkinase; 1.

```

DR SMART: SM00261; FU: 4.
 DR SMART: SM00219; TYRKC: 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP_1.
 DR PROSITE: PS00011; PROTEIN_KINASE_DOM_1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR_1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 25
 FT CHAIN 26 1308
 FT DOMAIN 26 651
 FT TRANSMEM 652 675
 FT DOMAIN 676 1308
 FT DOMAIN 186 334
 FT DOMAIN 496 633
 FT DOMAIN 718 985
 FT NP_BIND 724 732
 FT BINDING 751 751
 FT ACT_SITE 843 843
 FT DISULFID 189 197
 FT DISULFID 193 205
 FT DISULFID 213 221
 FT DISULFID 217 229
 FT DISULFID 230 238
 FT DISULFID 234 246
 FT DISULFID 249 258
 FT DISULFID 262 289
 FT DISULFID 293 304
 FT DISULFID 308 323
 FT DISULFID 326 330
 FT DISULFID 503 512
 FT DISULFID 507 520
 FT DISULFID 523 532
 FT DISULFID 536 552
 FT DISULFID 555 569
 FT DISULFID 559 577
 FT DISULFID 580 589
 FT DISULFID 593 614
 FT DISULFID 617 625
 FT DISULFID 621 633
 FT MOD_RES 162 162
 FT MOD_RES 168 1188
 FT MOD_RES 1258 1258
 FT MOD_RES 1284 1284
 FT CARBOHYD 138 138
 FT CARBOHYD 174 174
 FT CARBOHYD 253 253
 FT CARBOHYD 358 358
 FT CARBOHYD 410 410
 FT CARBOHYD 473 473
 FT CARBOHYD 495 495
 FT CARBOHYD 548 548
 FT CARBOHYD 576 576
 FT CARBOHYD 620 620
 FT CONFLICT 1062 1062
 FT CONFLICT 1080 1082
 SQ SEQUENCE 1308 AA; 146957 MW; D944B096A08B41 CRC64;

Query Match 43.2%; Score 2962; DB 1; Length 1308;
 Best Local Similarity 45.0%; Pred. No. 5.6e-15;
 Matches 609; Conservative 194; Mismatches 384; Indels 166; Gaps 29;

QY 1 MELA-ALCRWGLL--ALLPGAASVCTGTDMKRLPASPEHLMDLRLHLYGSCGVQ 57
 DB 1 MKLATGLWVGSLVAARTVPSASQSCAGTENTKSLSDLEQYBALRKRYENCCEVM 60

QY 58 GNLELTLYLPNTASLSFLDIOGVGVYLIANOVROVLORLATVRGTOULFEDYALAVL 117
 DB 61 GNLELTISIEHNDLSFLSIREVTGYVALNQFYPLENLRLIRGKLYEDYALAIIF 120

QY 116 DNGDPLNNTPTVTGASPGGLRELQURSLTEILKGVLIQRPOLCYQDPTILMKDIFKHKN 177
 DB 121 LNVKDKGNF-----GLQELGLKNLTELINGGVVYVDQNKFLCYADTIHQDITVNNPW 171

QY 178 QLATLIDITNRBRACHPSPMKSKSRGMSSESDCOSTRTVCAGGC-ARCKPLPTDCC 236
 DB 172 PSNMNTLVSTIGSSCGGCHSKSCIG-RWMTGETNHCQTLRTVCAEQDCKRCVPYVSDCC 230

QY 237 HEQCAAGCTGPKNSDCLACHFNHSGICEJHCALVYNTDTRESMPNREGRTFGASCY 296
 DB 221 HRECAAGSGGPKDIDCEACMNFDSGACVYQCQTFVYNTTQLHNFNPKAYTGAFCV 290

QY 297 TACRYNLTSTVGSCITLVCPHLNQEVAEDGTORCEKSPCARCVGGLGMFNFTVFW 356
 DB 291 KCPENFV-VDSSCVACPSKMEV-EENGMCKPCTDICKRACDGTG-----TGLS 342

QY 357 LRVKVSASHLEEFAGCKKI FGSIAFLPESPDPAISNTAPLOPEOLQVETLEITGYL 416
 DB 343 MSAGVDSNIDKFINCTKINGNLI FLVTGIDHDPYVAIDAIDPEKLVNRTVREITGFL 402

QY 417 YISAPDSLPPLSFQULQYIRGILHNGAVSLTGLGISMGLSLRSLGSLALHH 476
 DB 403 NIQTPPMVMDVSFVSNLVTTGGVLSGLLILKQGGITSLOFOSLKEISAGNIYTD 462

QY 477 NTHLCFVHTVPMQDLFNNPQALHTANRPEDECVGEGLAHQLCARHCWCGPTQCVN 536
 DB 463 NSNLCTYHTITWTLTFSTVNGRIYIRNRRANENCTAGMWCNHLCSGCGMGPGQCLS 522

QY 537 CSQPLRGCECEBQVULQGLPREVNAHCLPCHPEQCP-QNGSVTCFGEBAQVCACAH 595
 DB 523 CRFSRGKICIESCNLYDGEFERFENSGIIVECDSQCEKNEGDGLTCHGPPNCTKCSH 582

QY 596 YKDPFPCARCPGSKVCKDLSMPIMKFPDEGACQCPINCTHSCVLDLDDKGC----- 646
 DB 583 FKDGNCVCKEPDVLQGANF--LFKADDDRECHPHNCTQCGNPTSHDCIYYWTG 640

QY 649 ----PAQRASPLTSIYSAVY-GILVVVLGVVFGILIKRQCKIRKYTRRLLOETELV 703
 DB 641 HSTLPQHAR-TPL-IAAGVIGGLFVLIVALTFAVYVRKRSIK-KKRALRRL-ETELV 695

QY 704 EPLTSSGAMPQAMRLTEKTELAKVVLGSGAFGTYYKIMIPDGENNYKIPAIKYLE 763
 DB 696 EPLTSGAMPQAMRLTEKTELAKVVLGSGAFGTYYKIMIPDGENNYKIPAIKYLE 755

QY 764 NTSPEANKEILDEAVYVAGVSPVYSRLGLCTSTVQLVTLQMPYVGLDHYENRGL 823
 DB 756 TTGPANVEMFDEALIMASVDRPLVLLGVCCSPITQLVTLQMPHCCLEHYEHKNDI 815

QY 824 GSQDLINCMQIAQMSYLEDVRLVHNDLAARNVLYVSPHHVKITDGLARLLDDETEY 883
 DB 816 GSQDLINCMQIAQMSYLEDVRLVHNDLAARNVLYVSPHHVKITDGLARLLDDETEY 875

QY 884 HADGGKVPYIKMMALLESILRRFTQSDVMSYGVTVMEIMTFGAKPYDGI PAREIPOLLEK 943
 DB 876 NADGKMEIKMMALLECIIHYKFTQSDVMSYGVTVMEIMTFGAKPYDGI PTRREIPOLLEK 935

QY 944 GERLPQPICTITDYVMIMVKKWMI DSCRRPFRSLVSEFSRMARDPQRFVYIQNED-LGP 100
 DB 936 GERLPQPICTITDYVMIMVKKWMI DADSRPFKSLAEFSRMARDPQRFVYIQDDDKML 995

QY 1003 ASPLDSTFYRSLDDEDDMDGLVDAEYLVLPQGGFFCPDP----- 104
 DB 996 PSNDKSTFYRSLDDEDDMDGLVDAEYLVLP-QAENIPPLYTSTRITDSRSEIGHSPPP 105

QY 1042 --APGAGVHHRRSSSTRSGADLTGLPESEEA PRSPLAPSEBQADVDFDGLKMG 105
 DB 1055 AYTPMSGQFYVQDGFATQGG--MEMPTATSTIPEAPVA--QCATLEMFDDSCNG 110

QY 1100 AAKGIQSLPTHDPSPLQYRSFDPVLP-LS-----ETDGVAPLITSPPEYVQPDVR 115
 DB 1110 TURKRVVHVVEDSTQYSDPVPAPERNPRALDEBGMTPMDHKPQOEYLVNVE-- 116

QY 1153 POPSPPREGPLPAARPGATLERAKTUSPGKNGVAKVDVAFGGA VENVPEVLTFOGGAAPQ 121
 DB 1168 -----ENPVSNR-----KNQDLQ-----ALNDPEVHSAASG----- 119

QY 1213 PHPPPA-----FSPAFDNLYYWDODPPEBGA 123

DB 1195 --PRKAEDEVNPELYLNTFTNALGNAYMKNLSLSYPERAKKAFDNDYWNHSLPRST 1252
 QY 1239 --PSTFKGPT-----AENPEYL 1255
 DB 1253 LQHPDYLOEYSTKTFYKONGRIPIVAENPEYL 1285

RESULT 8
 XMRK_XIPMA STANDARD; PRT; 1167 AA.
 ID XMRK_XIPMA
 AC P13388
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
 GN XMRK OR TU.
 OS Xiphophorus maculatus (Southern platyfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorphi; Acanthopterygii; Percomorphi; Altheriomorphi;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OX NCBI_TaxID=8083;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90015140; PubMed=2797166;
 RA Witbrodt C., Adam D., Maitzschek B., Naeseler W., Raulf F.,
 RA Telling A., Robertson S.M., Scharl M.,
 RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
 RL inducing Tu locus in Xiphophorus.";
 RL Nature 341:415-421(1989).
 RN [2]
 RP REVISION TO 515.
 RA Scharl M.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 CC - FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC - SUBCELLULAR LOCATION: Type I membrane protein.
 CC - DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL, X16891; CAA34770.2; -.
 DR PIR, S06142; S06142.
 DR HSSP: P13362; 1FGK.
 DR InterPro: IPR000494; EGRF_L_domain.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00069; Kinasase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR PRINTS: PR00109; TRKINASE.
 DR ProDom: PD000001; Euk_Pkinase; 1.
 DR SMART; SMO0261; FU; 5.
 DR SMART; SMO0220; S_TKC; 1.
 DR SMART; SMO0219; TYKC; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 KM Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
 FT SIGNAL 1 25
 FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE

FT DOMAIN 26 642 KINASE.
 FT TRANSMEM 643 665 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 666 1167 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 710 977 PROTEIN KINASE.
 FT NP_BIND 716 724 ATP (BY SIMILARITY).
 FT BINDING 743 743 ATP (BY SIMILARITY).
 FT ACT_SITE 835 835 BY SIMILARITY.
 FT DISULFID 195 204 BY SIMILARITY.
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 220 228 BY SIMILARITY.
 FT DISULFID 224 236 BY SIMILARITY.
 FT DISULFID 237 245 BY SIMILARITY.
 FT DISULFID 241 253 BY SIMILARITY.
 FT DISULFID 256 265 BY SIMILARITY.
 FT DISULFID 269 296 BY SIMILARITY.
 FT DISULFID 300 311 BY SIMILARITY.
 FT DISULFID 315 330 BY SIMILARITY.
 FT DISULFID 333 337 BY SIMILARITY.
 FT DISULFID 504 513 BY SIMILARITY.
 FT DISULFID 508 521 BY SIMILARITY.
 FT DISULFID 524 533 BY SIMILARITY.
 FT DISULFID 537 553 BY SIMILARITY.
 FT DISULFID 556 569 BY SIMILARITY.
 FT DISULFID 560 577 BY SIMILARITY.
 FT DISULFID 593 615 BY SIMILARITY.
 FT DISULFID 618 626 BY SIMILARITY.
 FT DISULFID 622 634 BY SIMILARITY.
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1167 AA; 129934 MW; 4793EB0749DC1D55A CRC64;

Query Match 39.2%; Score 2688.5; DB 1; Length 1167;
 Best Local Similarity 45.0%; Pred. No. 2e-136;
 Matches 573; Conservative 171; Mismatches 385; Indels 145; Gaps 25

QY 4 AALRGWGLALLRPGASR---QVGTGDMKRLPASSETLDMRLHYOCQVQGN 59
 DB 8 AALLQ--LLVLISIRCCSIDPDRKVCQGSNQWTM--LDNHYLKVKKMYSGCNVLEN 62
 QY 60 LELTYLPTNASTFLQDIOEVQGYVLLAHQVQVPLQRLIYRGTLFEDNYALAVLDN 119
 DB 63 LEITYQENQDLSFLQIOEVQGYVLLAHNEVSTIPLVNLRLRQGLVHGNTLLVMSN 122
 QY 120 GDPINNTPTVTSAPGSGRLRELQSLREILKGYLIQRNQLCYDTILMKDIFHKKNQL 179
 DB 123 YQK-NPSSP--DVYQVGLKQLQSLNLEILISGVKXSHNPLCNVETIMWMDIVDKTSNP 179
 QY 180 ALTLIDFNRSPACHPSPMCKGSRGWSSESDCQSLRTVCAGGC-ARCGPLPTCCHE 238
 DB 180 TMLILPHAFERQCKCHGCGVMSGCA-PRGHQCKFKLLCAQCNRRCGPRPIQCNH 239
 QY 239 QCAAGCTGPKHSDGLACLHFNHSGICGLHCPALVTYNTDTFESMPNDEGRYFGASCVTA 298
 DB 240 HCAGGCTGPRATDCLACRDFNDQDTCOTPPRXIYDIVSHQVVDNPNIKYITGAACVXE 299
 QY 299 CPNYVLTSDVSGCTLVPLRNLNOEYTAAGQORCKSKPCARVCYGIQGMN-NFTVSFWL 357
 DB 300 CPSNIVYTE-GACVRSAGKLEVD-ENGRSKCPGCGVPRKVCDDGIGISLSNTTA--- 354
 QY 358 RVKVSASHLEEFAGCKKIFGSLAFLEPSFGDPASNTABLAPEOLQVPELLEITGYLY 417
 DB 355 ---VNSINISFNCIXINDIILNNSSEGGDPHYIGTMDEHLMNLTVAKEITGYLY 410
 QY 418 ISAMPDLSPLDSVQNQLVIRGLIHNGAYS-LTLQGLGISWGLRSLRELSSGLALIH 476

```

Db      411 IMMPEMNTSLSVQNLLEIRGRITTFGRGFSFVVVQVRHLOMLGRBLKSVSANGVILKN 470
Qy      477 NTHLCFVHTVMDQLFNNPHQALLHTANRPEDCEVGGACLCHQACARGHGWGGPPTQCN 536
Db      471 TLQIRYANTIMWRRLFSESDQSIYDART-----ENQTCNNECSDDGCGGPTMCVS 523
Qy      537 CSQFLRQGEVEECVLOGLPREVYNARHCLPCHPEQOPONGSVTCGPGPADQCVACAY 596
Db      524 CLHVDGRGRCVASCNLDGEPREAOVDGRCVQCHQELVQTDSTLCGPPANQSSAHF 583
Qy      597 KDPFCVARGCPGVKPDLSYMPIMKFPDEBGCQPCPINCTHSCVDLDKXCPAEOBPAP 656
Db      584 QDPQOCIRCPCHGILGDDDTL-IMKYADKMGCQPCPCNQNTCGGSGGSLGCGCD-IVSH 641
Qy      657 LHSIVASVGLLVVGLVGVFGGILIKRQCKIRKYMRRLQETELVEPLTPSGAMPNQ 716
Db      642 SSLAVGLVGLITVYALLIVLRRRLIK-RKRITRCLQELVEPLTPSGQAPNQ 700
Qy      717 QMRILKETELRKVYKLGSGAFGYKGIWIPDGENVXIPVAIKVLRENTSPKANKELDE 776
Db      701 FLRIKETEFKKDRVLGSGAFGYKGLMNDGENIRIPVAIKVLRATSFKVNOEVLDE 760
Qy      777 AYVAGVGSFYVSRLLGLTSTVQLTQMLPVGCLLDHRENGRGLSGODLLNMCQTA 836
Db      761 AYVAVSVDPHVCRLGLICTLSAVQLTQMLPVGCLLDYVQHGERLQGMWLLNMCVQTA 820
Qy      837 KQMSYLEDVRLVHRDLAARNLVKSPPNKVITDFGLARLADIBETEYHAGCKVPIKMA 896
Db      821 KGMNLEERHLVHRDLAARNLVKNPNVKITDGLSKLTLADKEVQAQAGGVPIKMA 880
Qy      897 LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIAPAEIPDLLEKGERLPORPICTD 956
Db      881 LESILQMTYTHQSDVMSYGVTVWELMTFGSKPYDGIAPAEIASYLENGERLPORPICTIE 940
Qy      957 VYMIWKCMMIDSECRPFRELYSEFSRMAPDOPFVIONEDIGPASPDSFTFYSLIE 1016
Db      941 YMIILKCMIDPSRRPFRELYVGFSGMARDPSRYVIOG--NLPSLSDRLLFSLLS 997
Qy      1017 DDDMGDLVDAEYVLPQGFPCPPAPAGAGMVHRRSSSTRGSGDLTLGLPSEEEA 1076
Db      998 SDD-DVDADEYVLLPYKRI-----NRQGS----- 1020
Qy      1077 PRSLAPSEAGSGSDVDFDGLDMGAKGLQSLPTHSPFLQRYSDPTV-PLPSETDGYVA 1135
Db      1021 -EPCLPTTGH-----PVENSITLRNISDPTQNALXKDLQGH-- 1056
Qy      1136 PLTCSPOPEYVNOPDVRPQ-----PSPRE-----GPP-AARPAQATLERAKTLSPG 1182
Db      1057 -----EYVNPQSETSRSLSDIYNNVYEDLTDGMPVSLSQEALTNFSRPEYLNTN 1108
Qy      1183 KKGVYKDVAFAGANENPEYILTPQGAAPQHPAPASPAFDNLYNDQPPERGAPPST 1242
Db      1109 QMSL--PLVSSGSDDDPY--QAG-----YQAA-----LPQTALGTGN 1143
Qy      1243 FKGTPTAENPEYL 1256
Db      1144 GMFLPAAENLEYLG 1157

```

```

OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90083234; PubMed=2687875;
RA      Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT      "Isolation and characterization of ERB3, a third member of the
RT      ERBB/epidermal growth factor receptor family: evidence for
RT      overexpression in a subset of human mammary tumors";
RL      Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90311312; PubMed=2164210;
RA      Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA      Todaro G.J., Shoyab M.;
RT      "Molecular cloning and expression of an additional epidermal growth
RT      factor receptor-related gene.";
RL      Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN      [3]
RP      SEQUENCE FROM N.A. (SHORT FORM).
RC      TISSUE=Placenta;
RX      MEDLINE=93282822; PubMed=7685162;
RA      Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT      "c-erb3 gene encodes secreted as well as transmembrane receptor
RT      tyrosine kinase.";
RL      Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC      -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULIN AND NTAK.
CC      -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC      (POTENTIAL).
CC      -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC      SECRETED (SHORT FORM).
CC      -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC      SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC      -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC      -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC      SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC      -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC      AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC      PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC      -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC      -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaborative
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M29366; AA35790.1; -
DR      EMBL; M34309; AA35790.1; -
DR      EMBL; S61953; AAB26935.1; -
DR      PIR; A36223; A36223.
DR      HSSP; P11362; IFGK.
DR      Genew; HGNC:3431; ERBB3.
DR      MIM; 150151; -
DR      InterPro; IPR000494; EGFR_L_domain.
DR      InterPro; IPR000719; Euk_DK_inase.
DR      InterPro; IPR002174; Furin-like.
DR      InterPro; IPR001245; Tyr_DK_inase.
DR      Pfam; PF00069; Pkinase; 1.
DR      Pfam; PF00757; Furin-like; 1.
DR      Pfam; PF01030; Recep_L_domain; 2.
DR      ProDom; PD000001; Euk_DK_inase; 1.
DR      SMART; SM00261; FU; 3.
DR      SMART; SM00219; TYRK; 1.
DR      PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR      PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
DR      PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR      Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
DR      Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;

```


DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbb-3 precursor (EC 2.7.1.112)
 DE (c-erbB3).
 GN ERBB3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RP [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RA MEDLINE=96096535; PubMed=8522190;
 RA Heller N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
 RT "Cloning of the rat ErbB3 cDNA and characterization of the
 RT recombinant protein";
 RL Gene 165:279-284(1995).
 RN (2)
 RP REVISIONS TO 85, 513 AND 565.
 RA Heller N.J., Koland J.G.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 922-1097 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
 RX MEDLINE=97184212; PubMed=9030624;
 RA Carroll S.L., Miller M.L., Frommet P.W., Kim S.S., Corbett J.A.;
 RT "Expression of neurogins and their putative receptors, ErbB2 and
 RT ErbB3, is induced during Wallerian degeneration.";
 RL J. Neurosci. 17:1642-1659(1997).
 CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = AMP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
 CC PHOSPHATIDYLINOSITOL 3-KINASE.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U29339; AAC28498.2; -
 DR EMBL: U52530; AAC3050.1; -
 DR HSSP: P11362; 1FGK.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00069; pkinase; 1
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recept_L_domain; 2.
 DR PRINTS: PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SMO0261; FU; 5.
 DR SMART: SMO0219; TYRK; 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; FALSE_NEG.
 DR PROSITE: PS00109; PROTEIN KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane: Glycoprotein; Multigene family; Receptor; Signal;
 FT SIGNAL 1 19
 FT CHAIN 20 1339
 FT DOMAIN 20 643
 FT EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 644 662
 FT DOMAIN 663 1339
 FT DOMAIN 183 259
 FT DOMAIN 707 964
 FT NP_BIND 713 721
 FT BINDING 740 740
 FT ACT_SITE 832 832
 FT DISULFID 186 194
 FT DISULFID 190 202
 FT DISULFID 210 218
 FT DISULFID 224 226
 FT DISULFID 227 235
 FT DISULFID 231 243
 FT DISULFID 246 255
 FT DISULFID 259 286
 FT DISULFID 290 301
 FT DISULFID 305 320
 FT DISULFID 323 327
 FT DISULFID 500 509
 FT DISULFID 504 517
 FT DISULFID 520 529
 FT DISULFID 533 549
 FT DISULFID 556 573
 FT DISULFID 576 585
 FT DISULFID 589 610
 FT DISULFID 613 621
 FT DISULFID 617 629
 FT CARBOHYD 126 126
 FT CARBOHYD 250 250
 FT CARBOHYD 353 353
 FT CARBOHYD 408 408
 FT CARBOHYD 414 414
 FT CARBOHYD 437 437
 FT CARBOHYD 469 469
 FT CARBOHYD 522 522
 FT CARBOHYD 566 566
 FT CARBOHYD 616 616
 FT CONFLICT 1028 1028
 SQ SEQUENCE 1359 AA; 147545 MW; 0AA5F2402BDF0FIE CRC64;
 Query Match 34.5%; Score 2365.5; DB 1; Length 1339;
 Best Local Similarity 40.7%; Pred. No. 4,36-119;
 Matches 525; Conservative 174; Mismatches 429; Indels 161; Gaps 34
 QY 3 LAALRMGLLALPFGAA---STQVCTGDMKRLPASETHLDKLRHYOGGVVQGN 59
 DB 7 LQVLC-----FLSLARGEEMNSQAVCGTLNGSLVGADNOVOTLYKYEKEVVMGN 62
 QY 60 LELTFLPTNAGSLFQDIOEVQGVLLAHQVROVPLQRLRIVRGTQLFEDNYALAVLDN 115
 DB 63 LEIVTGHNADLSFLQWIRVETGVVIVANMFSLPLPNLRVVRGTQVYGGKFAIFVM-- 120
 QY 120 GDPLNNTTPTVGASPGGRLQLRLSLEILKGVLIQRNPOLCYQDTILMKDIFHKKNQL 175
 DB 121 ---LNYNT-----NSSHALRLQKLFQTLTEILSGYVIEKNDKLCHMDIIDRDLTVR- 170
 QY 180 ALTLIDTRSRACHPCSPMKSGSKCWSESSSDQSLRTVCAAGC-ARCKGPLPTDCHE 236
 DB 171 GAETVYKNGANGCPRCHEVCKG-RCKGPGPDQDILTKTICAPQCNKRCGPPNQCHD 225
 QY 239 QCAAGCTPKKASDCLALHPNHSICELHCPALVYNTDFFESMPNDEGRYFGASCVTA 296
 DB 230 ECAGGCSGPQDTCFACRPFNSGACVRCPEPLVYKTLFQLEPNHRTKYQGVGVAS 285
 QY 299 CPYNTLSTDVSCITLVGLHNQETALDQGRCKKSPCARCYGICMNNNTVFSFWLR 358
 DB 290 CPNNV-VQGFPCYRACPDDMEVD-KGGLKMCPPCGGLCPKACEGGSGS-----R 339
 QY 359 VPKVASHLSEFAQCKKIFGSLAPSPFQDPAASNTAPLQPEGLQVFTLEBTIGLYI 418
 DB 340 YQTVDSNIDPFVNCTKILGNLDLTLGLNVDPNPKIPLADPELVNFRVREITGYLNI 399
 QY 419 SAMPSLPLDLSVFNQLQVIRIRIHNAYS-LTIGLGISWLGRLSRLGSLALIHNN 477

```

Db      400 QSWPHMNFNSVFNLTITGRSLYNKGFSLIMKNLNTVSLFRSLKEISARVYISAN 459
Qy      478 TILCFVHTVMDQLFRNPHQALLHTA-NRPEDECEVGEGLACHOLCARGMWGSPQVCN 536
Db      460 QQLCYHSHLNTRLRLRGSEERLDIKYDRPLGEBLAGKVCDDLSSGGGGMGSPQCLIS 519
Qy      537 CSQFLRGCEVCEERVLQGLPREVYNAHCLPCHPECPQONGSVTCGFPADQVCAAHY 596
Db      520 CRNYSREGVCVTHCNFLOGSEPREFVHAQCFSCPECLPMEGTSTCGSGSDACACAHF 579
Qy      597 KDPFCVAPRCPSPGKPLDLYMPKPKPDEEGACOPCINTGSC--VDLDDKCPRAEGR 654
Db      580 RDGHCVNNSCHGILG--AKSPIKYDAQNECRPCHEINTGCGNPELDDCGQAEVLM 637
Qy      655 SPLTIVSAVVGILLVVLGVGFGILIKRQOKIR-KYTRRLLOETELVEPLTPSGAMP 713
Db      638 SKPHLVIAVWG--LAVILMLGSGFLYMRGRIRIONKRAMRYLNGESIEFLDPS-EKA 694
Qy      714 NOAOMRLKETELKRVYLGAGACTYKGIWIPDGENVXIPVAIKYLRNTSPKANKEL 773
Db      695 NKVLARIFKETELKRLKVLGSGVGVHKGIMLPEGESIKIPVCIVIEDKSGRQSFQAV 754
Qy      774 LDEAVYVAGVSPYVSLGICLTSTVQLVTLMPYGLDHYRNRGRIGSGLLNMCM 833
Db      755 TDHMLAVGSLDHAIVRLGLCPGSSQLVTVYPLGSLLDHVKQHRITGLPOLLNMGV 814
Qy      834 QIAGMSYLEVRLVHNDLARNLVSPNHVKITDGLARLLDIDETEVHADGKVPK 893
Db      815 QIAGMYLLEHSHVHDLARNMLKSPQOVADGVADLLPPDDKQLHSEATPIK 874
Qy      894 WMALESILRRRTQSDVMSYGVTVMLMTFGAKPYDGIPIAREIPDLKEGERLPPQPC 953
Db      875 WMALESIHFGKITQSDVMSYGVTVMLMTFGAPYGLALIPDLKEGERLPPQPC 934
Qy      954 TIDVYMTVWKCMWIDSCRRPPELVSESPKARDPQFVYVIONEDLGPASPLDSTFYS 1013
Db      935 TIDVYMTVWKCMWIDENIRPFKELANEFTRMADPRPVIVIKRAS-GPQTP--PAERS 991
Qy      1014 LLEDDMDGLVDAEYLVPOGFCPPAPAGAGVWHHRSSSTSGGDLTLGIEPE 1073
Db      992 VLTREL-----QEALEPEL-----DLDLDEAE 1017
Qy      1074 E-----EAPRSLAPSEG-----AGSDVPDGLGCAKAGQ 1105
Db      1018 EGLATLSGALSILPTGLTRPRGQSILSSGVMYNGSLGEALCLDSVILGRRQFER 1077
Qy      1106 SLPHDPSPLORYEDETVPLPSTDGV---APL-----TC-----SPOPE---Y 1145
Db      1078 PLSLH-PIPRGR-----PASESEGHVGVSEAEIQEKVSVCRSRSRSPRPGDSAY 1129
Qy      1146 VNCQDVVRPOPSPRECP-----LPAAPACATLERAKTLSP-GKNGV----- 1187
Db      1130 HSGRHSLLTPVTPISPPGLEEDGNGVYMDTLRKASSREGTSLSVGLSTLGEED 1189
Qy      1188 KDVFAPGAVENPPLYLTPQGAAPQPHP 1216
Db      1190 ED-----EYEVYMKRRGRSP-PRFP 1209

```

```

OC      Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX      NCBI_taxid=7227;
RN      [1]
RP      SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX      MEDLINE=94350209; PubMed=8070664;
RA      Clifford R., Schubach T.;
RT      "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT      that several genetically defined classes of alleles cluster in
RT      subdomains of the receptor protein.";
RL      Genetics 137:531-550(1994).
RN      [2]
RP      REVISIONS.
RA      Clifford R., Schubach T.;
RL      Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
RX      MEDLINE=85124611; PubMed=2982499;
RA      Lynne E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT      "The Drosophila EGF receptor gene homolog: conservation of both
RT      hormone binding and kinase domains.";
RL      Cell 40:599-607(1985).
RN      [4]
RP      SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC      STRAIN=Oregon-R; TISSUE=Embryo;
RX      MEDLINE=87002474; PubMed=3093080;
RA      Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT      "Alternative 5' exons and tissue-specific expression of the
RT      Drosophila EGF receptor homolog transcripts.";
RL      Cell 46:1091-1101(1986).
RN      [5]
RP      SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP      ANALYSIS.
RX      MEDLINE=99102120; PubMed=9892502;
RA      Leeshkin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT      "Several levels of EGF receptor signaling during photoreceptor
RT      specification in wild-type, Ellipse, and null mutant Drosophila.";
RL      Dev. Biol. 205:129-144(1999).
RN      [6]
RP      SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC      STRAIN=Berkley;
RX      MEDLINE=20196006; PubMed=10731132;
RA      Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA      Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA      George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA      Suton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA      Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA      Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA      Abail J.F., Agbayani A., An H.-J., Andrews-Piankoff C., Baldwin D.,
RA      Ballev R.M., Basu A., Baxendale J., Bayraktoglu L., Beasley E.M.,
RA      Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA      Botkova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
RA      Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA      Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA      de Pablos K., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA      Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA      Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA      Foster C., Gabrielian A.E., Garg N.S., Gehlert W.M., Glaser K.,
RA      Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA      Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA      Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA      Jalali M., Kalush F., Kappen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
RA      Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA      Liakou P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA      Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA      Merkulov G., Milshina N.V., Mobarry C., Morris J., Mostreli A.,
RA      Mount S.M., Moy M., Murphy B., Murphy L., Murty D.M., Nelson D.L.,
RA      Nelson S.R., Nelson K.A., Nixon K., Nusser D.R., Pacheco J.M.,
RA      Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA      Reinartz K., Remington K., Saunders R.D.C., Schejter F., Shen H.,
RA      Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA      Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA      Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA      Wang Z.-Y., Waasman D.A., Weinstock G.M., Weissbach J.,

```

RA Williams S.M., Woodage T., Morley K.C., Mu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of *Drosophila melanogaster*";
RL Science 287:2185-2195(2000).
[7]
RP SEQUENCE OF 959-1078 FROM N.A.
RC STRAIN=Daekwanyeeong;
RX MEDLINE=85137938; PubMed=2983232;
RA Madsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RT "A *Drosophila* genomic sequence with homology to human epidermal
RL growth factor receptor.";
RL Nature 314:178-180(1985).
[8]
RP SEQUENCE OF 1133-1137, 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
RP ANALYSIS.
RX MEDLINE=92038942; PubMed=1936959;
RA Raz E., Schejter E.D., Shilo B.Z.;
RT "Intercellular complementation among DER/flb alleles: implications for
RL the mechanism of signal transduction by receptor-tyrosine kinases.";
RL Genetics 129:191-201(1991).
[9]
RX REVIEW.
RA MEDLINE=97248481; PubMed=9094709;
RA Perrimon N., Perkins L.A.;
RT "There must be 50 ways to rule the signal: the case of the *Drosophila*
RL EGF receptor";
RL Cell 89:13-16(1997).
[10]
CC -1- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GIREN, VTN AND ARGOS,
CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
CC ESTABLISHMENT OF VENTRAL CELL FATES. MAINTENANCE OF ANTIOSCOSA
CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
CC CUTICLE.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
CC PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: TYPE I (SHOWN HERE), TYPE II AND
CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: UBICUOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLGEN OF
CC TESTIS AND OVARY AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
CC AND THORACIC AND ABDOMINAL GANGLIA.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AF052754; AAC08536.1; -;
DR EMBL, AF052753; AAC08536.1; JOINED.
DR EMBL, AF052754; AAC08535.1; -;
DR EMBL, AF052752; AAC08535.1; JOINED.
DR EMBL, K03054; AAA51462.1; -;
DR EMBL, K03417; AAA51460.1; -;
DR EMBL, K03416; AAA50965.1; -;
DR EMBL, K03418; AAA51461.1; -;
DR EMBL, AF109077; AAD26134.1; -;
DR EMBL, AF109078; AAD26132.1; -;
DR

[illegible]

```

Db 423 IDGNIRIDQTFSGPODVYANTYMTGPRYIPLDPERREVFSTVKEITGYLIEGTHPOFRN 482
Qy 428 LSVFONLOVITGRILIHNGAY-SLTLQIGISWLCRLRSIREIGSLALIHNTLCPHNT 486
Db 483 LSYFNLETITGRQLMESMFAALAIKVSLSYLSLMRYLKQISSSVIYQNRDLCTVSN 542
Qy 487 PMDOLFNRPHQALLHTANRPEDECVGEGGLACHQLCARHCGMPGPTQCVNCSQFLRGQEC 546
Db 543 RMPAIQKEPEQKQVWVNNENLRADLCEKNGTICSDQCNEDGCGAGTDCITCKNFENGTG 602
Qy 547 VEECRVLOGLEPREYVNAHCLPCHPEQCPONGSTYTCGPEPDQCVAAHAKDPFPCARC 606
Db 603 IADCGYISNAK--FDNRCKIHPBECR-----TCGAGADHCQECVHRDQHCSEC 654
Qy 607 P-----SGVK-----PDL 614
Db 655 PNKNYNDRGVCRCHATGCGTGPDTIGACTTCNLAIINNDATYKRCCLKDKDPD- 713
Qy 615 SYMPIWK--PDEGACQ-----CP-----NCH----- 638
Db 714 GY--FMEYVHPQEQSLKPLAGRAVCRKCHPLCLCTNYGHEQVCSKTHYKRRQECET 771
Qy 639 -----SC-----VDLDKG----- 647
Db 772 ECPADHYTDEBQECQFQHPRECNGCTGAGDDCKSCNFKLPDANETGYVNSMTNCTS 831
Qy 648 -CPAQR-----ASPLTS-----IVSAVGIILVVYGVVFGI 679
Db 832 KCPLEMRHVNQYTAIGPYCAASPPRSKKTANLDVNMIFITGAVLVPTICLCV--T 889
Qy 680 LIKROOKIRKYT--MRLLQETELVEPLTPSGAMPQOMRIKTELAKVAVLSSGAF 757
Db 890 YICROKQAKKETVYMTALSGCEDSEPLRSNIGANLCKRIYKDLKELKGGVLGGAG 949
Qy 738 GTVYKGINIPDGENVKLPVAIKVLRENTSPYANKELIDEAVYVAGVSPYVSLGICLT 797
Db 950 GRVYGVWVPEGENVKIPVAIKELKSTGARSSEFLREAVYMASEHVALTLKLVACS 1009
Qy 798 STVCLVTQLMPYGLLDHVRNKGRLSGODLLNMCQIAGKMSLELVRLVYHDLAARNV 857
Db 1010 SQMMLITQLMPLGCLLDYVNNRNRKISGKALLNSTQIAGKMSLEKRLVHDLAARNV 1069
Qy 858 LVKSPNHYKITDFGLARLLIDETEYHADGKVPDKMMALLESILRRFTQSQVWSYGT 917
Db 1070 LVQTPSLVKIITDFGLAKLLSDSNRYKAAGKMTKMLALECIRNRYFTSKDYMAFGVT 1129
Qy 918 VWELMTFGAKPYDGIPAEIPDLLEKGERLPQPICTIDVYMWIMVCKMIDSECRPRE 977
Db 1130 IMELTTFQORPHENIPAKDIPDLIEVGIKLEQPEICGLDIYCTLSCWHLDAAMRPTFKQ 1189
Qy 978 LVSEFSMAADPOPFVVIQNEDLG--PASPLDSTFFYSLLDD--DMGDLVAEEYIVP 1032
Db 1180 LTVFAEFARDPGRYLAIPGDKFRLRA-----YTSQEKDLIRKLAFTTGSSALAK 1242
Qy 1033 QQGFCDPDPAPGAGGMVHRRSSSTRSGGDLTLGLPSEEEAP-----RSPLAPSEG 1086
Db 1243 PDDYLQPKYAPGPS-----HRTDCT-----DEMFKINRYCROPSKNS 1281
Qy 1087 AGSDVFDG---DLGMGAAGKGLQSLPTHDPSELYQYSDPTVPLPSETDGVVALTLTSPQP 1143
Db 1282 TGDDERSDAREVGVGNIR-----LDLPVEDDDVLM--TQCPGP 1319
Qy 1287 AGSDVFDG---DLGMGAAGKGLQSLPTHDPSELYQYSDPTVPLPSETDGVVALTLTSPQP 1143
Db 1144 EYVNPQDVPRQPPSPREGPLPAARPAATLERAKTLSPGKGVGVDFAGGAVENPEYL 1203
Db 1320 NNNNNNN-----NPNQNNMAAVGVAAGYM-----DLIGVPVSDNPEYL 1358
Qy 1204 -----TPQGAAPQR-----PPAFSP-APDNLTYMD 1230
Db 1359 LNAQTLGVGESPIPTQITGIPVWGGPQTMVKVPMPSSEPTSSDHEYND 1408

```

RESULT 12

```

ERBB_ALV
ID ERBB_ALV STANDARD; PR: 634 AA.
AC P00534;
DT 21-JUL-1998 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nielsen T.W., Matoney P.A., Goodwin R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-D.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor."
RL Cell 41:719-726 (1985).
CC -| CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -| MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INFECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -| MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -| SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaborative
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC
CC EMBL, M1006; AAA48763.1; ALT_INIT.
DR PIR, A00643; TVCHLV.
DR PIR, B00643; TVFVLV.
DR HSSP, P11362; IFCK.
DR InterPro, IPR000719; Euk_Pkinase.
DR InterPro, IPR001245; Tyr_Pkinase.
DR Pfam, PF00069; Pkinase.1.
DR PRINTS, PR00109; TYRKINASE.
DR ProDom, PD000001; Euk_Pkinase.1.
DR SMART, SM00219; Tyrc; 1.
DR PROSITE, PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE, PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE, PS50011; PROTEIN_KINASE_DOM; 1.
DR Transferrase; tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

Query Match 25.5%; Score 1749.5; DB 1; Length 634;
Best Local Similarity 52.3%; Pred. No. 1.5e-86;
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17

Qy 593 CAHKKDPFCYARSPSSVKEDLSMPYKRPDEGACQPPINICTHSCVDLDDKGPAPQ 652
Db 3 CAHRTDPHCYKACPAVIGENDTL-VMKYADANAVCOLCHPNCITGCKGPGLEGCP--- 58
Qy 653 RASPLTIVSAV-VGLVVVYLVGVVFGILIKRQOKIRKYTMRLLOETELVPLTPSGA 711
Db 59 NGSKTPSIAGVGVGGLCLVVGIGIGLYARR-HIVRKTLRLQOEELVPLTPSGE 117
Qy 712 MENOQMRILKTELKRVKVLGSGAGFQYVYKGIWPDGENVKIPVAIKVLRNTSPKANK 771

```

```

Db      118 APNQHRLILKETEFKVKVLGSGAFGVYGLWIMPEGEKXKIIVAIKELREATSPRANK 177
Qy      772 EILDEAYVMAGVSGPYVSRLLIGICTSTVQVLTQMLPGCLLDHVRENRGRLSQDLLNM 831
Db      178 EILDEAYVMASVDNPHVCRLLIGICTSTVQVLTQMLPGCLLDHVRENRGRLSQDLLNM 237
Qy      832 CMQIAKMSYLEDEVLRVLDLAARVLRVLSFNHYKTFDGLARLLIDDETEYHADGGKVP 891
Db      238 CVQIAKGMVYLEERLVRDLAARVLRVLTQHKITDPSGAKLGDDEKYEHAEGSKVP 297
Qy      892 IKMMALESILRRRFTHSDVWSYGVTTWELMTFGAKPYDGIIPAREIPDLLKGERLPPOP 951
Db      298 IKMMALESILHRIYTHSDVWSYGVTTWELMTFGSKPYDGIIPASEISSVLEKGERLPPOP 357
Qy      952 ICTIDVYIMVWKCMWIDSECRPRERELVSFESRMARPPQRFVVIQ-NEDIGPASPDLSTF 1010
Db      358 ICTIDVYIMVWKCMWIDSECRPRERELVSFESRMARPPQRFVVIQ-NEDEGASPLDSTF 417
Qy      1011 YRSLLIEDDMDGLVDAEYLVPOQGFPCPPAPGAGGVHRRHRSSTRSGGDLTLGLE 1070
Db      418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy      1071 PSEEFAPRSP-----APSEGAGSVDFDGLMGAKGLQSLPTHDSPLQRYSEDPYV 1125
Db      450 -----SRTPLLSSLSATSNNSATNCID-----RKQGHFVREDSFVQRYSSDPTGN 495
Qy      1126 LPSET--DGYVAPLTCSPQPEYVNPQVPCPPSPREGPLPAARPAATLERAKTUSPGK 1183
Db      496 FLEESIDOGFL-----PAPEYVNG--LMPKKPS-----TAMVQ 526
Qy      1184 NGVAVKDV-----AFGAVENPEVLTPOGCAAPPHPPAPSPAFDNLTY 1228
Db      527 NQIYNNISLTAISKLPMDSRYSNSTAVDNPEYL-----NTNOSPAAKTGVFESSPY 578
Qy      1229 WDQ-----DPEP-----RGAPSTFGKGTPTAENPEYLGDPV 1260
Db      579 WIGSGNHQINLDNPDYQODFLPNEIKPKGLKLVPAENPEYLRVAAP 625

RESULT 13
ERBB AVIER STANDARD; PRT; 604 AA.
ID ERBB AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (rel. 01, last sequence update)
DT 21-JUL-1986 (rel. 01, last sequence update)
DE 15-UN-2002 (rel. 41, last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroviruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbb gene of avian erythroblastosis virus is a member of the src
RT gene family."
RL Cell 35:71-78(1983).
RP
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RA Debire B., Henry C., Benalassa M., Bisette G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erba gene of avian erythroblastosis virus reveals a
RT new type of oncogene."
RL Science 224:1456-1459(1984).
RC -; CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC -; tyrosine phosphate.
CC -; DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.

```

```

CC -; MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -; SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaborative
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1; -.
DR PIR; A06644; TVYOH.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; Pkinase; 1.
DR ProDom; PD000001; Tyrc; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50111; PROTEIN_KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R->W (IN REF. 2).
FT CONFLICT 140 140 S->F (IN REF. 2).
FT CONFLICT 146 146 I->V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76EBDD06745D609 CRC64;

Query Match 24.9%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 4,4e-84;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16

```

Db	CAHRIDGPHCVKACPRAGVLGENDTL-VKRYADANAVQCLCHPNCRCGKGGLSGCP----	5
Qy	653 RASLTSTVAVV-GILLVVLGVGFGLIKRQOKIRKTYRRLQSTELVELVETPSGA	711
Db	59 NGSTPPIAGVGVGGLCLLVVQLGIGLYLRSR-HIVKRTLRLQLQERLEVEPTSGE	117
Qy	712 MPNOAKRIKETELRKYVKSAGATGYKGIWIPDGENVYIPAIIVLEBNTSPYANK	772
Db	118 APNOAHLRIKETEPFKYKVLGFGAGCTYKGLMIPRGEKTIIPALIELDEAISPANK	177
Qy	772 EILDEAVYMAVGSFYVSRLLGICLTSTVQLVLTQIMPGCLLDHYENRGRSLGSDLLNW	831
Db	178 EILDEAVYMAVSNPHVCRLLGLCTISTVQLITQIMPGCLLDYIREKHONIGSYLLNW	23
Qy	832 CMQIAKMSVLEDEVRLVHRDLAARNVLYNSPNNHYITTPQGLARLLDIETETVHADGGVYR	897
Db	238 CVQIAKMSVLEERHWHVHRDLAARNVLYVTPQHVITTPQGLAKQUGADEKEYHAEGGVYR	297
Qy	892 IKMMALLESIRREFRTHQSDVWSYGYVMEIMTFGAKFYDGIIPAREIPDLLEKGERLPQRP	953
Db	298 IKMMALLESILHRIYTHQSDVWSYGYVMEIMTFGSKPYDGIIPASISVLEKGERLPQRP	35
Qy	952 ICTIDVTYVWKKCMKMDISECRPRFELVSEFSRMAKRDQRFVYIO-NEDLCPSAPLDSFT	1017
Db	358 ICTIDVTYVWKKCMMSDADSRPKFRELILAEFSKMAKRDPRRYLVIGDGRMKLPSPTDSKF	417
Qy	1011 YRSLLEDDMDGDLVDVAEYLVPOQGFCECDPAPAGAGWVHHRHSSYRSGCGULTGLE	1071
Db	418 YRILMEEDMEDIVDADEVYLVPRQGF-----NSPST-----	444
Qy	1071 PSEBEARSPUL-----APSEGAQSDVFDGDLGMAKGLQSLFTHDPSPQLQYSEDPVP	1124
Db	450 -----SRTPLLSLSATSNKATNIDANG-----H-----	476
Qy	1126 LPSTDSYVAPPLCSQPEYVNOQDVRPPSPRFGRLPAARPAQAT-LEPAKTLSECKN	1185
Db	477 -PVREDDFL-----PAPRYVNO-LMPKPSPTAMVONQIYNYISLTAISKLPIDSRYON	527
Qy	1185 GVAQDVAFPGAVENPEYL 1203	
Db	528 -----SHSTAVDNEPYL 539	
RESULT 15		
EGFR, CHICK		
AC	EGFR, CHICK STANDARD: PRT, 703 AA.	
AC	P1387:	
DT	01-JAN-1990 (Rel. 13, Created)	
DT	01-JAN-1990 (Rel. 13, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER)	
DE	(Fragment).	
GN	EGFR.	
OS	Gallus gallus (Chicken).	
OC	Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi;	
OC	Archosauria, Aves, Neognathae, Galliformes, Phasianidae, Phasianinae;	
OC	Gallus.	
OX	NCBI_TaxID=9031;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=88261272; PubMed=3260329;	
RA	Lak I., Johnson A., Howk R., Sarp J., Bellot F., Winkler M.,	
RA	Ullrich A., Vennstrom B., Schlessinger J., Givol D.,	
RT	"Chicken epidermal growth factor (EGF) receptor: cDNA cloning,	
RT	expression in mouse cells, and differential binding of EGF and	
RT	transforming growth factor alpha."	
RL	Mol. Cell. Biol. 8:1970-1978 (1988).	
CC	-1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,	
CC	AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND	
CC	VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).	
CC	-1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein	
CC	tyrosine phosphate.	
CC	-1- SUBCELLULAR LOCATION: TYPE I membrane protein	

CC	-1-	MISCELLANEOUS: Binding of EGF to the receptor leads to	
CC	CC	dimerization, internalization of the EGF-receptor complex.	
CC	CC	induction of the tyrosine kinase activity, stimulation of cell DNA	
CC	CC	synthesis, and cell proliferation.	
CC	-1-	SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.	
CC	CC	-----	
CC	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	CC	use by non-profit institutions as long as its content is in no way	
CC	CC	modified and this statement is not removed. Usage by and for commercial	
CC	CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/	
CC	CC	or send an email to license@isb-sib.ch).	
CC	CC	-----	
DR	DR	EMBL: M20386; AAA48760.1; -	
DR	DR	InterPro: IPR000494; EGFR_L_domain.	
DR	DR	InterPro: IPR000719; Euk_Pkinase.	
DR	DR	InterPro: IPR002174; Furin-like.	
DR	DR	InterPro: IPR001245; Tyr_pkinase.	
DR	DR	Pfam: PF00757; Furin-like; 1.	
DR	DR	Pfam: PF01030; Recep_L_domain; 2.	
DR	DR	SMART: SM00261; Fy; 4.	
DR	DR	PROSITE: PS00107; PROTEIN_KINASE_ATP; PARTIAL.	
DR	DR	PROSITE: PS00109; PROTEIN_KINASE_TYR; PARTIAL.	
DR	DR	PROSITE: PS50011; PROTEIN_KINASE_DOM; PARTIAL.	
DR	DR	Transmembrane; Glycoprotein; Receptor; Signal; Transferase;	
KW	KW	Tyrosine-protein kinase; ATP-binding; Phosphorylation.	
FT	FT	SIGNAL	1 30
FT	FT	CHAIN	31 >703
FT	FT	DOMAIN	31 654
FT	FT	TRANSMEM	655 667
FT	FT	DOMAIN	668 >703
FT	FT	DISULFID	197 206
FT	FT	DISULFID	201 214
FT	FT	DISULFID	222 230
FT	FT	DISULFID	226 238
FT	FT	DISULFID	247
FT	FT	DISULFID	243 255
FT	FT	DISULFID	258 267
FT	FT	DISULFID	271 298
FT	FT	DISULFID	302 314
FT	FT	DISULFID	318 333
FT	FT	DISULFID	336 340
FT	FT	DISULFID	517
FT	FT	DISULFID	519 530
FT	FT	DISULFID	533 542
FT	FT	DISULFID	546 562
FT	FT	DISULFID	565 581
FT	FT	DISULFID	569 589
FT	FT	DISULFID	592 601
FT	FT	DISULFID	605 627
FT	FT	DISULFID	630 638
FT	FT	DISULFID	634 646
FT	FT	CARBOHYD	134 134
FT	FT	CARBOHYD	190 190
FT	FT	CARBOHYD	200 200
FT	FT	CARBOHYD	359 359
FT	FT	CARBOHYD	368 368
FT	FT	CARBOHYD	420 420
FT	FT	CARBOHYD	573 573
FT	FT	CARBOHYD	578 578
FT	FT	CARBOHYD	613 613
FT	FT	CARBOHYD	633 633
FT	FT	CARBOHYD	648 648
FT	FT	NON_TER	703 703
Q	Q	SEQUENCE	703 AA; 77427 MW; AFE2DE1B735A690 CRC64;

Query Match 23.4%; Score 1600; DB 1; Length 703;
Best Local Similarity 44.0%; Pred. No. 1.7e-78;
Matches 314; Conservative 115; Mismatches 250; Indels 34; Gaps 13;

QY 8 RRGILLALIPGAA-----STOVCTGTMDKRLRPASPETHLMDLRHLYQGCGVQGNLE 61

D6	13	RGAAVLVLLILGVALCSAVEEKKVCQGTNNKLTQIGHVBDHFTSLQRMNNECVLNSLE	72
OY	62	LTIPLPMASLSFLODIOEVQGYLLIHNQYRQVPLQRLIIVGTQLFEFNPLALVLDND	12
D6	73	IIVYEHNRDITLFTKTIQEVAGVYLLIANNVDYIPLENDIIRGNVLYDMSFALAVLSNH	13
OY	122	PLNNITFVTSAPGSLREIQLRSUTEILKGVLIQENPOLCVODTILMKDIPIHKNNQLAL	18
D6	133	-MKTO-----GRLPMKRLSILLNGVYKISNNPPLCMNDYLANMDIIDTSK-PL	18
OY	182	TLID-TNRSPACHPCSPMCKGRKWCSESSDDCQSLRTVACGCA-RCXGPLETDCHEQ	23
D6	183	TVLDFASNLSSCKRCHPNCTEDHCWAGRONCOTLTKVICAQCCSGRCKXPSDDCHNQ	24
OY	240	CAAGCTGPKHSDDCLACHFNHSGICELHOPALVTYNTDTFESPMNEGRYTGASCVTAG	29
D6	243	CAAGCTGPRSDCLACRKKFPDDATCQDTPPLVLYNPPTYQMDVNEBGRKSGATVREBC	30
OY	300	PYNVLSTDVSGCTLYCPRLHNOETVAEDGTQRCCKSPCARVYCYGLGMENNFPSFWLRY	35
D6	303	PHNYVVDHSCSVRSCNTDTEYV-ENGVRKCKCKGCLSKVONCIG-----IGELKGI	35
OY	360	PKYSASLLEEFACKCKIFGSLAFLEPSFQGDASNAPLQPELDQVFELEETGLYLS	41
D6	360	LSINATNIDSFKNCTKINGDVSTLPAFAGDAFTKLPDPKKLDVFRVKEISGLLQ	41
OY	420	AMPDSLDELVFONLQVIRGRILHNAVSLTLOGLSIMGLRSLRELSGALLIHNHT	47
D6	416	AMPENATDVAFENLEIRKRTKHQHOYSLAVVNLKIQSLGURSLEKSIDGDIATMKNN	47
OY	480	LCFVHTVPMQDLFRNPHQALLHTANRPEDECVGEGSLACHOLCARGHCMGPRPTOCVNC	53
D6	476	LCVADTMMMSLEPATOSOKTKIIQNKNKNDCTADRYACDPLCSDVQCMOPGFHCPSPCR	53
OY	540	FLRQECVEECRYLOGPREYVVARHCPLCHPCCQFONG---SVTFGEBAOCVACARY	59
D6	536	FSRQKECYKCCNIIQGEPRFERDSKCLPHSCLVQNSTAYITTSGGGPRDHCKMAHF	59
OY	597	KDPEFCVACRPSGVKPDLSYMPIWKEPRDEGACQPCPINCTHSCVDLDDKGPRAEQRAS	65
D6	596	IDGHCHYKACRPAVLGENDTL-VMKADANAVACQLCHPNCCTGCKSGPGLGCP---NGSK	65
OY	657	LTSIVSAVV-GILLVVLGVVPRGILLIKRQQRKIKYTMRLLOEFTLVEPLRP	708
D6	652	TPSLAAAGVAGGLLCLVNVVGIGLULYRRR-HIVRKSTLRRLLOERLVLVPLRP	703

Search completed: July 22, 2003, 08:45:59
Job time : 21.2793 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.5887 Seconds

(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-325-339-14

Perfect score: 6852

Sequence: 1 MELALCRWGLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP_ARCHA:.*
2: SP_BACTERIA:.*
3: SP_FUNGI:.*
4: SP_HUMAN:.*
5: SP_INVERTEBRATE:.*
6: SP_MAMMAL:.*
7: SP_MHC:.*
8: SP_ORGANELLE:.*
9: SP_PHAGE:.*
10: SP_PLANT:.*
11: SP_TODENT:.*
12: SP_VIRUS:.*
13: SP_VERTEBRATE:.*
14: SP_UNCLASSIFIED:.*
15: SP_VIRUS:.*
16: SP_BACTERIAP:.*
17: SP_ARCHAEP:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6219	90.8	1259	6 O18735	O18735 canis famli
2	3150	46.0	1209	11 Q9GX70	Q9GX70 rattus norv
3	3121	45.5	1210	11 Q9EP98	Q9EP98 mus musculu
4	2723	33.7	1165	13 Q9YH40	Q9YH40 xiphophoru
5	2708.5	33.5	1137	13 Q9W6F6	Q9W6F6 gallus gall
6	2313	33.8	1328	13 P79754	P79754 fugu rubrip
7	2030.5	29.6	1433	5 Q9B1H9	Q9B1H9 anopheles g
8	1872.5	27.3	419	4 Q9UK79	Q9UK79 homo sapien
9	1739	25.4	367	11 Q8R2X1	Q8R2X1 mus musculu
10	1720	25.1	729	15 Q86712	Q86712 avian rous-
11	1718	25.1	567	15 Q86714	Q86714 avian rous-
12	1697.5	24.8	412	4 Q8WYV0	Q8WYV0 homo sapien
13	1660	24.2	962	15 Q64895	Q64895 avian eryth
14	1645	24.0	545	15 Q85468	Q85468 avian eryth
15	1512.5	22.1	655	11 Q9WVF5	Q9WVF5 mus musculu
16	1496.5	21.8	643	11 Q9ERV6	Q9ERV6 mus musculu

17	1275	18.6	1193	5 Q9Y1X8	Q9Y1X8 ephydatia f
18	1200.5	17.5	1368	5 Q23821	Q23821 caenorhabdi
19	1180	17.2	1717	5 Q26566	Q26566 schistosoma
20	1131	16.5	527	13 Q90836	Q90836 gallus gal
21	1007.5	14.7	478	11 Q9ESF0	Q9ESF0 rattus nor
22	947.5	13.8	599	13 Q9PSH2	Q9PSH2 gallus gal
23	906	12.2	165	4 Q14256	Q14256 homo sapien
24	887	11.9	176	11 Q923V5	Q923V5 rattus nor
25	806.5	11.8	346	13 P11776	P11776 xiphophoru
26	778	11.4	435	5 Q8SZM1	Q8SZM1 drosophila
27	754.5	11.0	311	13 Q99162	Q99162 xiphophoru
28	739.5	10.8	1362	13 Q9PV44	Q9PV44 xenopus la
29	734	10.7	331	4 Q9BUD7	Q9BUD7 homo sapien
30	723	10.6	149	6 Q9BGG6	Q9BGG6 oryctolagus
31	723	10.6	1671	5 Q9NTV5	Q9NTV5 biophthalari
32	694.5	10.1	1368	13 Q8UW85	Q8UW85 paralicth
33	689	10.1	1418	13 Q93457	Q93457 scophthalm
34	678.5	9.9	1369	13 Q8UW86	Q8UW86 paralicth
35	666.5	9.7	1472	5 Q9US58	Q9US58 demdix mori
36	666	9.7	1412	13 Q8UW84	Q8UW84 paralicth
37	659	9.6	1358	13 Q73798	Q73798 xenopus la
38	640.5	9.3	1418	13 Q8UW83	Q8UW83 paralicth
39	636	9.3	1245	13 Q9YGH8	Q9YGH8 scophthalm
40	626.5	9.1	2144	5 Q9VD94	Q9VD94 drosophila
41	621	9.1	1371	11 Q9QWT4	Q9QWT4 rattus sp.
42	601	8.8	987	11 Q91YU0	Q91YU0 mus musculu
43	598	8.7	987	11 Q99WK2	Q99WK2 homo sapien
44	592	8.6	935	4 Q96L35	Q96L35 homo sapien
45	587.5	8.6	1036	4 Q07912	Q07912 homo sapien

ALIGNMENTS

RESULT 1
O18735 PRELIMINARY; PRT: 1259 AA.
ID O18735
AC O18735;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JUN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Erdb-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT cDNA cloning of erdb-2 from canine mammary gland.;
RT Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB008451; BAA2127.1; -;
DR HSSP: P1362; IFGX.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000494; EGFR_L domain.
DR InterPro: IPR000719; Euk_Dkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_kinase.
DR InterPro: IPR004019; YFP_motif.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; Kinase; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR Pfam: PF02757; YFP_2.
DR Prodom: PD000001; Euk_kinase; 1.
DR SMART: SM00261; Fy; 3.
DR SMART: SM00219; Tyrc; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

Query Match 90.8%; Score 6219; DB 6; Length 1259;
 Best Local Similarity 90.7%; Pred. No. 0;
 Matches 1149; Conservative 46; Mismatches 58; Indels 14; Gaps 4;

QY 1 MELAALCRWMLLALLPFGAASVQVCTGDMKRLBPASPELHMLRHLVGGCCVYVQNL 60
 1 MELAAMCRWMLLALLPFGAAGVQVCTGDMKRLBPASPELHMLRHLVGGCCVYVQNL 60
 DB 61 ELTYLPTNASTLFLQDIQEVQGVYLTAHQVQVPLQRLIRVQGLPQEDNATLAVLNG 120
 QY 61 ELTYLPTNASTLFLQDIQEVQGVYLTAHQVQVPLQRLIRVQGLPQEDNATLAVLNG 120
 DB 61 ELTYLPTNASTLFLQDIQEVQGVYLTAHQVQVPLQRLIRVQGLPQEDNATLAVLNG 120
 QY 121 DPLNNTVTVTASPGGLRELQRLSLTEILKGGVYLQGNPOLCYDTLWKDIFHNQOLA 180
 DB 121 DPLNNTVTVTASPGGLRELQRLSLTEILKGGVYLQGNPOLCYDTLWKDIFHNQOLA 180
 QY 181 LTLIDTRSRACHPCSPMCKSGRCWGESEDCSLTRTVCAAGCARCKGPTDCCHQC 240
 DB 181 LTLIDTRSRACHPCSPMCKSGRCWGESEDCSLTRTVCAAGCARCKGPTDCCHQC 240
 QY 241 AAGCTGKHSDCIACLHPNHSIGICELHCPALVYNTDTESPNPEGRYTFGASCVTACP 300
 DB 241 AAGCTGKHSDCIACLHPNHSIGICELHCPALVYNTDTESPNPEGRYTFGASCVTACP 300
 QY 301 YNYLSTDVGSCTVCPLEHNOVTAEDGTQRCCKSCPKARVQVYGLGMPNFTVSMRL-V 359
 DB 301 YNYLSTDVGSCTVCPLEHNOVTAEDGTQRCCKSCPKARVQVYGLGMPNFTVSMRL-V 359
 QY 360 PKVSASHLEEFACCKIFGSLAFLPESFDGPASNTAPAPQEPLOVFELEETGYLIS 419
 DB 360 PKVSASHLEEFACCKIFGSLAFLPESFDGPASNTAPAPQEPLOVFELEETGYLIS 419
 QY 420 AMPDLSPLDSVFQNLQVIRIRLHNQAYSLTLOGLSLWGLRSRLRELQSLALIHNTH 479
 DB 420 AMPDLSPLDSVFQNLQVIRIRLHNQAYSLTLOGLSLWGLRSRLRELQSLALIHNTH 479
 QY 440 AMPDSLPLSLSVFQNLQVIRIRLHNQAYSLTLOGLSLWGLRSRLRELQSLALIHNTH 479
 DB 440 AMPDSLPLSLSVFQNLQVIRIRLHNQAYSLTLOGLSLWGLRSRLRELQSLALIHNTH 479
 QY 480 LCFVHTVPMQDLFRNPHQALHTANRPEDECVGEGLAQCLCARHCWGPPTQCVNSQ 539
 DB 480 LCFVHTVPMQDLFRNPHQALHTANRPEDECVGEGLAQCLCARHCWGPPTQCVNSQ 539
 QY 474 LCFVHTVPMQDLFRNPHQALHTANRPEDECVGEGLAQCLCARHCWGPPTQCVNSQ 532
 DB 474 LCFVHTVPMQDLFRNPHQALHTANRPEDECVGEGLAQCLCARHCWGPPTQCVNSQ 532
 QY 540 FLRGQECVCECRYLQGLPREVYVNAHRLCPHPECOPONGSVTCFPEADQCVACHYKP 599
 DB 540 FLRGQECVCECRYLQGLPREVYVNAHRLCPHPECOPONGSVTCFPEADQCVACHYKP 599
 QY 533 FLRGQECVCECRYLQGLPREVYVNAHRLCPHPECOPONGSVTCFPEADQCVACHYKP 592
 DB 533 FLRGQECVCECRYLQGLPREVYVNAHRLCPHPECOPONGSVTCFPEADQCVACHYKP 592
 QY 600 PFCVARGSPGVKPDLSYMPIWKEPDEEGACOECPINCHSCVDLDDKCPAQRASPLTS 659
 DB 600 PFCVARGSPGVKPDLSYMPIWKEPDEEGACOECPINCHSCVDLDDKCPAQRASPLTS 659
 QY 593 PFCVARGSPGVKPDLSYMPIWKEPDEEGACOECPINCHSCVDLDDKCPAQRASPLTS 652
 DB 593 PFCVARGSPGVKPDLSYMPIWKEPDEEGACOECPINCHSCVDLDDKCPAQRASPLTS 652
 QY 660 IVSAVGIILVVVLGVVFGILIKRQOKIKRYTMRLLQETELVEPLTPSGAMPNQAQMR 719
 DB 660 IVSAVGIILVVVLGVVFGILIKRQOKIKRYTMRLLQETELVEPLTPSGAMPNQAQMR 719
 QY 653 IIAAVVGIILAAVVVGLVILIKRQOKIKRYTMRLLQETELVEPLTPSGAMPNQAQMR 712
 DB 653 IIAAVVGIILAAVVVGLVILIKRQOKIKRYTMRLLQETELVEPLTPSGAMPNQAQMR 712
 QY 720 ILKETELRKVKVLSGSAFGTVYKGIWIPGENVKIPVALKVRRENTSPANKKILDEYV 779
 DB 720 ILKETELRKVKVLSGSAFGTVYKGIWIPGENVKIPVALKVRRENTSPANKKILDEYV 779
 QY 713 ILKETELRKVKVLSGSAFGTVYKGIWIPGENVKIPVALKVRRENTSPANKKILDEYV 772
 DB 713 ILKETELRKVKVLSGSAFGTVYKGIWIPGENVKIPVALKVRRENTSPANKKILDEYV 772
 QY 780 MAGVGSFYVSRLLGICLTSTVQVLTQMLPEYGLLDHVENRGRLSODLWMCQIAKGM 839
 DB 780 MAGVGSFYVSRLLGICLTSTVQVLTQMLPEYGLLDHVENRGRLSODLWMCQIAKGM 839
 QY 773 MAGVGSFYVSRLLGICLTSTVQVLTQMLPEYGLLDHVENRGRLSODLWMCQIAKGM 832
 DB 773 MAGVGSFYVSRLLGICLTSTVQVLTQMLPEYGLLDHVENRGRLSODLWMCQIAKGM 832
 QY 840 SYLEDVRLVHRDLAANNVLYKSNHXTDFGLARLLDIDETRYADGKVPDKMALES 899
 DB 840 SYLEDVRLVHRDLAANNVLYKSNHXTDFGLARLLDIDETRYADGKVPDKMALES 899
 QY 833 SYLEDVRLVHRDLAANNVLYKSNHXTDFGLARLLDIDETRYADGKVPDKMALES 892
 DB 833 SYLEDVRLVHRDLAANNVLYKSNHXTDFGLARLLDIDETRYADGKVPDKMALES 892
 QY 900 ILRRFTHQSDVMSYGVWELMTFAKPYDGIIPAREIDLEKEGRLPPOPICITIDVYM 959
 DB 900 ILRRFTHQSDVMSYGVWELMTFAKPYDGIIPAREIDLEKEGRLPPOPICITIDVYM 959
 QY 893 IIPRRFTHQSDVMSYGVWELMTFAKPYDGIIPAREIDLEKEGRLPPOPICITIDVYM 952
 DB 893 IIPRRFTHQSDVMSYGVWELMTFAKPYDGIIPAREIDLEKEGRLPPOPICITIDVYM 952
 QY 960 IMYKCMWIDSECPRRRELVSFSRMAKDPQSFVVIQNEIDLQASPLDSTFYRSLDEDD 1019
 DB 960 IMYKCMWIDSECPRRRELVSFSRMAKDPQSFVVIQNEIDLQASPLDSTFYRSLDEDD 1019
 QY 953 IMYKCMWIDSECPRRRELVSFSRMAKDPQSFVVIQNEIDLQASPLDSTFYRSLDEDD 1012
 DB 953 IMYKCMWIDSECPRRRELVSFSRMAKDPQSFVVIQNEIDLQASPLDSTFYRSLDEDD 1012
 QY 1020 MGDLDVAEEYLVQGGFPCGDPAPAGAGVWHHRSSSTRSGGDLTLGLPSESEARSR 1079
 DB 1020 MGDLDVAEEYLVQGGFPCGDPAPAGAGVWHHRSSSTRSGGDLTLGLPSESEARSR 1079

DB 1013 MGDLDVAEEYLVQGGFPCGDPAPAGAGVWHHRSSSTRSGGDLTLGLPSESEARSR 1079
 QY 1080 PLAPSSGASDVFDGDLGMAKAGIQSLPTHDPRLQRYSEPTVPLPSESDGYAPLTC 11
 DB 1073 PLAPSSGASDVFDGDLGMAKAGIQSLPTHDPRLQRYSEPTVPLPSESDGYAPLTC 11
 QY 1140 SPQEPVYVQDPVAPPPSPREGPLPAARPAATLBER-----AKTLPKNGVYKDVAFAG 11
 DB 1133 SPQEPVYVQDPVAPPPSPREGPLPAARPAATLBER-----AKTLPKNGVYKDVAFAG 11
 QY 1195 GAVENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPERGAAPSTFKGPTAENPEY 12
 DB 1193 GAVENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPERGAAPSTFKGPTAENPEY 12
 QY 1255 LGLDVVPV 1261
 DB 1253 LGLDVVPV 1259

RESULT 2
 09QX70
 ID 09QX70 PRELIMINARY; PRT: 1209 AA.
 AC 09QX70;
 DT 01-MAY-2000 (TEMBLrel. 13, Created)
 DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
 DE Epidermal growth factor receptor.
 GN EGFR.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RX MEDLINE=90258888; PubMed=2342466;
 RA Petch L.A.; Harris J.; Raymond V.W.; Blaeband A.J.; Lee D.C.;
 RA Earp H.S.;
 RT "A truncated, secreted form of the epidermal growth factor receptor is
 RT encoded by an alternatively spliced transcript in normal rat tissue."
 RL Mol. Cell. Biol. 10:2973-2982(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RA Petch L.A.;
 RL Submitted (NOV-1991) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RA Guttridge K.; Dawson T.L.; Earp H.S.;
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
 RX EMBL: M37394; AAF14008.1; -;
 DR HSSP; P11362; IFGK.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF01030; Recep_L domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU_3
 DR SMART; SM00219; TYRK; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00101; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KM ATP-binding; Receptor; Tyrosine-protein kinase.
 SQ SEQUENCE 1209 AA; 134891 MW; 96FE7F6CC1B773 CRC64;

Query Match 46.0%; Score 3150; DB 11; Length 1209;
 Best Local Similarity 50.1%; Pred. No. 5e-229;

Matches .644; Conservative 169; Mismatches 348; Indels 124; Gaps 29;

```

QY      3 LAALGRWGLLLALBPGA-ASTOVCTGTDMKRLPASPETHLDMLRHYQCGQVQCNLE 61
      15 LAALGAG-----GALBEKKVCGTSTNRLTQGTGFEDHFLSQHFNNECEVVLNLE 66
QY      62 LTYLPTNASLSLQDOIQEVQGVLLAHNQVROVPLQRLRIYRGTOLEFDNYALAVDNGD 121
      67 IYVQRNDLSFLKTOEVAGYVLLALNTVERIPLNLIIRGNALVENTYALAVLSN-- 124
QY      122 PLNNTPTVGAAPGARELQSLRSLTEIKGVLIOQNPQLCYODTIIMKDIFFHKNQAL 181
      125 -----YGTNKTGRLRPMNLQELIGAVRENNPILCMETIQKRDIV-ODVLSN 175
QY      182 TLIDTNRS-RACHPSPCMKSGSRCSSESSEDCQSLTRTVCAAGCA-RCKGPLPTDCHEQ 239
      176 MSMDVQRHLTGCPKDPSCPCNGSCWGRGEENCQKLTIKCAQCSRCRCGRSPSDCHNQ 235
QY      240 CAAGCTGPRHSDCLACIHNHSGICELHCPALVTYNTDTFESMENPEGRYFGASCYTAC 299
      236 CAAGCTGPRHSDCLACIHNHSGICELHCPALVTYNTDTFESMENPEGRYFGASCYTAC 295
QY      300 PNYVLTVDVGSCTLVCPHNDQEVTAEDGTORCEKSKPCAVCYGLGM--FNNFTVSEFWL 357
      296 PNNYVVTDHGSCVRAAGPDYEV-EDGVSKCKCKDGPCKVCKNGIGIGEPKD-TLS--- 350
QY      358 RVPKVSASHLEFPACCKTIFGSLAPLESFDDPASNAP:QPEQLQVFEITLEITGLY 417
      351 ---INATIKHFKYCTALISGLHLPLPAFKGDSFTRPDPPELELTKVKEITPFL 406
QY      418 ISAMPDSLPLSVFONLQVIRGRILHNGAVSLTLOGLSIMLGRSLRELSGALTLHHN 477
      407 IQAMPENMTDLHAFENLEIRGRTKQHOFSLAVAGLNTLSGLRSLKEISDGVITISGN 466
QY      478 TFLCEVHTVPMQDLFRNPHQALLHTANPEDECEGEGIACHQLCARGHGWCWPGPTQVC 537
      467 RNLQYANTIMMKKLTGTNOKTKIMNNRAEKDCATNHCNPLCSSSEGCWPEPTDVCSC 526
QY      538 SQFLRGQECVEECRVLOGLPREYVNAHGLCPHPCQCPQNSVYCFGEHLDQCACAHYK 597
      527 QNVSRGRECVDCNILEGEPRFEVNSCICQHEBCLPQTNNITTCGTGPNCKICAHYV 566
QY      598 DPFPCVACPSGCVKBDLSYMPIWKPDEGACOPPCINCHSCVDLDDKGPAPORASP- 656
      587 DGHCVKTCPSGIMGNNTL-VMKFADANNVCHLCHACTYCGAGPGLKGC--QQPSPK 643
QY      657 LTSISAVVIGILLVVLGVNPGI-LIKRQOKIKRYTKRLLQETTELVEPLTPSGAMPNQ 715
      644 IPSIATGIVGGLFTIV-VALGIGLFMRERQLVRKRTLRLLQERREIVEPLTPSGEAPNQ 702
QY      716 AQMRILKETELRKVAVLGSAGAFYVKGIMTDPGENYKIPAVIIVLENTSPKXNKETLD 775
      703 AHLRLKETEFKIKIVLGSAGAFYVKGIMTDPGENYKIPAVIIVLENTSPKXNKETLD 762
QY      776 EAYVAVAGVSPYVSLGLICTSTYQVLTQIMPPYGLLDHYRENRGLSGQDLINCMQI 835
      763 EAYVAVASVDPHVCCLLGLICTSTYQVLTQIMPPYGLLDHYRENRGLSGQDLINCMQI 822
QY      836 AKGNSYLEDEVRLVHNDLAARVNVKSPHVKITDPGLARLLDIDETVYHADGGGVPIKX 895
      823 AKGNSYLEDEVRLVHNDLAARVNVKSPHVKITDPGLARLLDIDETVYHADGGGVPIKX 882
QY      896 ALESILRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIDPLEKGRLLPQPICTI 955
      883 ALESILRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIDPLEKGRLLPQPICTI 942
QY      956 DYYMIMVCKWMI DSCRRPRELISESRMARDQRFVIO-NDLDGAPAPLDSTFIRSL 1014
      943 DYYMIMVCKWMI DSCRRPRELISESRMARDQRFVIO-NDLDGAPAPLDSTFIRSL 1002
QY      1015 LEDDDMGDLVAEEYLVQOGFCGDPAPAGAGVWHRRHSSSTRSGGDLTLGLEPSEE 1074
      1003 MEEEDMEDVDADEVLIPOGGF-----NSPST----- 1030

```

```

QY      1075 EAPRSELPASGEGSDVDPDGLGMGAQGLSLPETHDPSPIQRYSEDPVLPSET--DG 11
      1031 --SRTPLLSLLSANSEN-----SSTVACINRNSCRVKEDAFIQRSSDPTSVLTEDNIDT 10
QY      1133 YVAPLTCSPQPEYVNOQDPVPPQPPSPREPGELPAAPAGATLERAKTISPGKGVKDYFA 11
      1085 FL-----FVPEYINQ-SVPRKPAQSVGNPYHNOPLHP-----APGDLHYQN--P 11
QY      1193 FCGAVENPEYL-TPOGGAAPQHPPEPAPFADNLVYWDQ-----DP-----PE 12
      1128 HSNAYSNPEYILNTAQ-----PTCISSGFDSALMIQSGHQMSLDNDYQODFPFK 11
QY      1236 KGAPSTPEKGPETAENPEYLGADV 1260
      1179 EAKPNGIFKQ-PTAENAEYLRVAP 1202
QY      1179 EAKPNGIFKQ-PTAENAEYLRVAP 1202

```

RESULT 3

```

Q9EP98 ID Q9EP98 PRELIMINARY; PRT; 1210 AA.
AC Q9EP98;
DT 01-MAR-2001 (TRMBLrel. 16, Created)
DT 01-MAR-2001 (TRMBLrel. 16, last sequence update)
DT 01-JUN-2002 (TRMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor isoform 1.
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTA;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maibhe N.U.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms."
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Schehl C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maibhe N.U.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms."
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.

```

DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR ATP-binding; Receptor; transferase.
 SQ SEQUENCE 1210 AA; 134840 MW; 62CD021C9DB32E18 CRC64;

Query Match 45.5%; Score 3121; DB 11; Length 1210;
 Best Local Similarity 49.6%; Pred. No. 7.8e-227;
 Matches 635; Conservative 171; Mismatches 354; Indels 120; Gaps 26;

QY 11 LLIALLPPGAA-STGYCTGTMLKRLPASPTHLDMRLHYQCGQVQVQNLLETLYPTN 68
 14 LITLCAAGALEEKVCQSTNRLTQLGTFEDHFLSQRMNNEVALLNLETITVQRN 73
 QY 69 ASLSFLQDIQVQVYLIANQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 128
 74 YDLSFLKIQEVQVYLIANQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 124
 QY 129 VTGASPGLELEQLRSLEILKGVLIQENPOLCYQDTILWKDI----FKKNQALATLI 184
 125 -YGTNRTGLRELMRLQELIGAVFNNPILCNMDITQWRMDIQVNFVSNMSMDL--- 180
 QY 185 DTKRSACPCSCMCKGSCMGSESSDQSLRTVCAGCA-RCKGRLPTCCGCGCAAG 243
 181 -QSHPSCKPCDPSGNGSCMGSESCQKLTIIQAQCSHRCKGRSPDCCNQCAAG 239
 QY 244 CTGPKHSDCLACLFHNSGICELHCPALVTYNTDPESMPNDEGRYFGASCYTAAPNY 303
 240 CTGPRSDCLVCCQFQDEATCKDTCPLMLYNTTYQMDVNEGKYSFGATCVKCKPNY 299
 QY 304 LSTDVSCILVCLAHQVETAEDEGTORCEKCKPCARVCYIGM-FNNFTVSWLRVPK 361
 300 VMTDHSQVRAQCPDYEV-EEDGIRCKCKCGPRKVCNGGIGQEFKD-TLS----- 350
 QY 362 VSASHLEFPAGCKKIRIGSLAFLPESFDGDPASNTAPLQPEQLOVETLEETLYISAM 421
 351 INKTNKHKYCALSIGDHLIPVAFKDSFRTPLDRELEIKTKETITGELLQAM 410
 QY 422 PSLPPLSPQNTQVIRGRILHNGAVSLTQGLGISWLSRLSGLALHNNHLC 481
 411 PDNMTDLHFEENLEIRGRKQHGQSLAVVGLNTSLGLSLKEISGDVYISGNRLC 470
 QY 482 FVHTVPMQLFRPHQALHTANRPEDECVSEGLACHQLCARHGQMGSPQCNQOFL 541
 471 YANTINWKKLFGTPNCKTKIMNNRAEKDAVNVHVCNPLCSSEGCGWGPEDCVSCQNV 530
 QY 542 RQGEVCEGRVLOGLPREVYNARHCLPCHPECOQNGSVTCFGEPAQCVACAYKDPF 601
 531 RGRECEKCNILEGEPRFVENSECICHPCLPQANNITCTGRGPDNCCICAHYIDGPH 590
 QY 602 CVARCSGVKPLSLYPIKFPDEBACQPCINCTHSCVDLDKGCAPAEORASPLSIV 661
 591 CVKTCGAGIMGENNTL-VKRYADANNVCHLCHANCTYGAGAGLQGCWESGKIRISIA 649
 QY 662 SAVVGLLVVGLGVFI-LIKRQCKIRKYMRLLOETELVEPLTSGAMPNQAQRI 720
 650 TGIVGGLFIVV-VALIGLFMRHRHIVKRLRLLOERELVEPLTSGAPQAQAHRI 708
 QY 721 LKTELKRVKVLQSGAFVYKGINIPDGENYKIVALKVLENTSPANKELLIDEAYVM 780
 709 LKETEPRKIKVAGSGAFVYKGLMPEBEKVKIVALKVLENTSPANKELLIDEAYVM 768
 QY 781 AGVGSYVSRLLGICLITVQVLTQMLPGCLLDHVRNENRGLSGDPLNKCQIAGMS 840
 769 ASVDNPHVCRLLGICLITVQVLTQMLPGCLLDHVRNENRGLSGDPLNKCQIAGMS 828
 QY 841 YLEADVLRVRLAARNVLYKSNVYKIDFGIARLLDIDETFYHADGKVPKIMMALESI 900
 829 YLEDRVLVHRDLAARNVLYKTPQHYKIDFGIARLLDIDETFYHADGKVPKIMMALESI 888
 QY 901 LRERFTHOSDWSYGVTVELMTFGAKPYDGIPIREIPDLLEKGRRLPQPICTIDVYMI 960

DB 889 LHRITHQSDVWSYGVTVELMTFGSKPEYDGIPIASDISILEKGERLPPQPICTIDVYMI 94
 QY 961 MKCKMIDSECPFRRELVSERSSRAARPPQFVYIQ-NEDGASPLDSTYRLEDD 10
 949 MKCKMIDSDSPKRELILESXKARPPQFVYIQGERHMLSPFDSNYRALMEED 10
 QY 1020 MGDVLDAEYVLPQCGFCPCDPAPGAGMVRHRSSTRSGGDLTLGLEPSEEAAPS 10
 1009 MEDVDADADLYLPQCGF-----NSPST-----SRT 10
 DB 1080 PLAPSEGASDVFDGDLGNAKGIQSLPTHPDSEFLQYSDPTVPLPSET--DGYAPL 11
 1035 PLISISATSN-----NSTVACINRNGSCRYVEDAFLLCYSSDPGAVTEDNIDAFI--- 10
 QY 1138 TCSPOEYVNOQDVPVPPSPREGLPAPAPAGATLERAKTLSPGKGVKDVFAFGAV 11
 1088 ---PPEYVNO-SVKRPAQSVQNFVYHNOPLR-----AFGRDLHYON--PHNAV 11
 DB 1198 ENPEYL-TPQGAAPQPPPPAFSPADNLYWDO-----DF-----BERGAP 12
 1134 GNPBYLNTAQ-----FTLSGFGNSPALMTQKSHQMSLDNPDVQDPFPEKTKPN 11
 QY 1241 STFKGPTAENPEYLGIDVP 1260
 DB 1185 GIFKG-PTAENAYLRVAP 1203

RESULT 4
 Q9YH40
 AC Q9YH40; PRELIMINARY; PRT; 1165 AA.
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Receptor tyrosine kinase proto-oncogene.
 GN XMRK
 OS Xiphophorus xiphidium.
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 RN [1]_Taxid=8066;
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIO PURIFICATION;
 RX MEDLINE=98241172; PubMed=9582016;
 RA Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Dueschl J.,
 RT "Activation of the Xmrk proto-oncogene of Xiphophorus by
 overexpression and mutational alterations.",
 RL Oncogene 16:1681-1690(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIO PURIFICATION;
 RA Schartl M.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U53471; AAD10500.2; -
 DR HSSP: P11362; 1FGK.
 DR InterPro: IPR000345; CytC_heme_bnd.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Gram_pos_anchor.
 DR InterPro: IPR001245; Ty_Pkinase.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR PRINTS: PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; Fy_3.
 DR SMART: SM00261; TyKc; 1.
 DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN 1.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN 1.

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 2.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Kinase; Tyrosine-protein kinase.
 SQ SEQUENCE 1165 AA; 129614 MW; 7f7E838D871A74E CRC64;

Query Match 39.7%; Score 2723; DB 13; Length 1165;
 Best Local Similarity 45.3%; Pred. No. 9, 1e-197;
 Matches 580; Conservative 172; Mismatches 379; Indels 148; Gaps 31;

QY 1 MELALCRWGLLALPPG-AAST-----QVCTGDMKRLPASPETHLDMRHLYGCGV 55
 Db 4 LELLEL-----LILLILSIGCCSTDPDRVCQGTGNTMT--LDNHLKKKKMYSGCNV 56
 QY 56 VQGNLELTPTNPAASFLQDIQEVGYLLAHNQVRYPLQRLATVNGTOLFEDNATA 115
 Db 57 VLENLEITTYQENQDLSFQSIQEVGGYLLAHNVSTPLVNNLIRGQNLVYGNFTLL 116
 QY 116 VLNDGDLNNTTPTVGASPGRLRELQLSLTEILKGYLIQRNPLQCYQDTITLKKDIFHK 175
 Db 117 VMSNYOK-NPSS--DYVQGLKQQLSLTEILSGYKVSHPNPLCNVETINMMDLYDK 173
 QY 176 NNQALTLIDINRSRACHPGSPCKSGSRCSGSSSDCQSLTRTVCAAGC-ARCKGPLPTD 234
 Db 174 TSNPTMNLIPHAFFERQCKCDPCGVNCSGMAFPQHCQKFTKLCAEQCNRRCRGPKPID 233
 QY 235 CCECCAGAGCTGPKKSLACLPHNHSICELHCPALVTYNTDIFESMPNDEGRTFGAS 294
 Db 234 CCEHCAGAGCTGPRADCLACRDFDDGCTCKTCEPKITVIVSHQVVDNNINITYTGAA 293
 QY 295 CVTACFPYNTLSDVGSCTLVCPAHQVEYTAEDGTORCEKCKSPCARVCYGLGMEN-NFTV 353
 Db 294 CYKECPSNVYTE-GACVRSAGAGLEV-D-ENGKRSCKPCQGVCPKVDGIGISLSNTI 351
 QY 354 SWMLRPKYASASHEFPAGCKKIFGSLAFLESPFGDASTAPLQEPQLQVFTLEIT 413
 Db 352 A-----VNSTINISFSNCTKINDIILNRNFEFGDHYXIGMPDEHMLNLTVEIT 404
 QY 414 GYLISAMPDSLPLDSVFONLQVIRGRILHNGAYS-LTLQGLISWLGRLRELISGLA 472
 Db 405 GYLIVMMPENNTLSLVFONLEIIRGRITFGSGSFVYVQVSHQWGLSLKVSAGNV 464
 QY 473 LTHNTHLCFHTVPWDQLFRNFQALHLHTANRPEDCEVGEGLACHOLCAQHMGWGPPT 532
 Db 465 LKNTPOLRYASTIMWRLLFRSEDSQSIYDART-----ENQCNNECSHDCGWPSPPT 517
 QY 533 QCVNCSQFLRGQECVEEGRYQGLPREVYNAHCLPCHPEQPNQSGTCGEPADQVA 592
 Db 518 MCVSLHYDRGRCVASCNLLQGEPRERAOVGRVCQHQECLVQTDLSLTCYGPAPANCSK 577
 QY 593 CAHYDPPFCVAPRCPSGVKPDLSYPIKFPDEEGACQCPINCTHSCVDLDDKGCPEAQ 652
 Db 578 CAHPDGPQCIIRCPHGMIGDGTI-IRKYADKMGQCPHQNTQCGSGSGGLSGCGD- 635
 QY 653 RASPLSTISAVVGLLVVYLVGVGILIKRQCKIRKTYMRLLQETELVEPLTPSGAM 712
 Db 636 IVSHSLAVGLVSGLLITVIALLVLLVLRRLIK-RKRTIRRLQSEKIVEPLTPSGQA 694
 QY 713 PNOQMRLIKETELRKVYLGSGAGFYKGIWIPDGENVXIPAIIVLRBNTPSKANKE 772
 Db 695 PQOAFRLIKETEPFKDKRVLGSGAGITVYKGLNMPDGENIRIIPALIVLEASPKNQE 754
 QY 773 TLDEAYVMAVGSPPVSLGLCLTSTVQVLTQMLPYGCLLDHVRNGSLGSDLLNMC 832
 Db 755 VLDEAYVMAVDHPVCRLLGLCLTSAVQVLTQMLPYGCLLDYVRQHERICQGMILNMC 814
 QY 833 MOIAGMSYLEPVRIVHRDLAARNVLYVSPHYKITPGLARLIDIDETVYHDDGKVPPI 892
 Db 815 VOIAGMNYLEPVRIVHRDLAARNVLYVSPHYKITPGLSKLTADDEKTYQAHGKVPPI 874
 QY 893 KMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGRLLPQPI 952
 Db 875 KMALESILQMTYTHQSDVMSYGVTVWELMTFGSKPYDGIIPAKIASVLENGERLQPI 934

QY 953 CTIDVYMVKCMIDISECRPRFRELVEFSRMAADPQRFVVIQNEMLGPASPLDSTFYR 10
 Db 935 CTIEVYMILKCMIDPSSRPREFRELVEFSQMAADPQRFVVIQ--NLSPSPDRILFS 99
 QY 1013 SLLEDDMDKLDVDAEAYLVPOGFCFPPAPAGAGMVRHRSSSTRGCGDILGLEFS 10
 Db 992 RLSSDD--DVVDADEYLL-----RYKRIN-RQGS-----10
 QY 1073 EEEAPRPLAPSEGAGSDVFDGLGMAKGLQSLPTHPDPLQRYSEDPTV-PLPSETD 11
 Db 1019 -----EPICPENGH-----PVENSIALRYISDPQNALERKLD 10
 QY 1132 GYVAPLTSPQPEYVNPQVPEQP-----PSPRE-----GLP-AAAPGATLERAKT 11
 Db 1053 GH-----EYVNOGSETSRSLSDIYNPNEDLTDGWPVSLSSQEAETNFSRPEY 11
 QY 1179 LSPKNGVYKDVFAFGVAVENREYLTPOGGAAPQHPAPASPPAFDNLYYDQDPEBGA 12
 Db 1103 LMTNONSLL--PLVSSGSMDDPY--OAG-----YQAAF-----LPQGA 11
 QY 1239 PPSRFGKTPFAENPEYGL 1257
 Db 1138 LTGNGWFLPAENLEYGL 1156

RESULT 5

QY 096F6 PRELIMINARY; PRT; 1137 AA.
 AC 096F6;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Receptor Tyrosine Kinase (Fragment).
 GN ERB34.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=HINDBRAIN;
 RX MEDLINE=99263203; PubMed=10328884;
 RA Dixon M., Lumsden A.,
 RT "Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in
 RT embryonic chick hindbrain.";
 RL Mol. Cell. Neurosci. 13:237-258 (1999).
 DR EMBL, AF121963; AAD31764.1; -.
 DR HSSP; P11362; 1FGK.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR000719; Euk_Pkinase.
 DR InterPro; IPR002174; Euk_Pkinase.
 DR InterPro; IPR001368; TNFR_C6.
 DR InterPro; IPR001245; Tyr_Pkinase.
 DR InterPro; IPR004019; YLP motif.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF01030; Recept_L_domain; 1.
 DR Pfam; PF02157; YLP; 2.
 DR SMART; SM00261; FU; 3.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_Pkinase; 1.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.
 KW Kinase; Tyrosine-protein kinase.
 FT NON TER 1
 SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;

Query Match 39.5%; Score 2708.5; DB 13; Length 1137;


```

QY 302 NYLSTDVSGCTLVCPHNOEYTAEDGTOR-CEKSKPQCARCYGLGNFNNFTVSWLRLVP 360
DB 290 HFV-VDGSSCVSVCPDPMEV--ERGSROCELCGSCPKKCEGTGAERQOT----- 338
QY 361 KVSASHLEFPAGCKKIFGSLAFIPESFPGDPAASNTAPLOPELOVEFTLEIGLYISA 420
DB 339 -VDSNINISFINCTKIQSILHFLVTGILGDPKQVPPIDAKKLEVPFVLEIDILNIQS 397
QY 421 WPDSLPDLVSFQNTQVIRGRILHNGAVSLTQGLGISWGLSRSLRELSGSAIJIHNTHL 480
DB 398 WPKELNDLSVFSLSLTITIGRSLEFKRFSIMVMRIPILNISGLRSLSREISDSVVISQNAHL 457
QY 481 CFVTVPMWDLFRNH-QALLHTANRPDECEVSGSLACHQICAGHGWGPGPOCVNCSQ 539
DB 458 CYHHTVMTQLEFRGSRVANSLSNRPAECVAGRGVCDPLCSGSGWGPGPOCCLSRN 517
QY 540 FLRGOEVEECVVLQGLPREYVNAH-CLPCHPECPONGSVTCFGEAQCACAHYKD 598
DB 518 YSRHOTCVAGCHFNNGIIPREFAGLVGVACHPECKPQTGA SCTGGADECACTKFRD 577
QY 599 PPFVACRCPGKPLSTMPYWKPPDEGACOPCPINCHSCVDLDKCGPAEORASPLT 658
DB 578 GPYCMSSCPAGVN-DEKGLIFKEPNRGHCEPHONCTQCGSGPGLNDC---LEAARLT 633
QY 659 STVSAVVGLLVVVGVEF-----GILIRROQKIRKYMRLQETELVEPLTPSGA 711
DB 634 ISSGOITGIALCVPAGLIFCLVLPFLGMLYRGLAIRKRAMRYLESSEFELPG-GE 692
QY 712 MPNOQMRILKTELKRYKVLGSGAFGTGKGIIPDGENYKIPVAIKVRENTSPKANK 771
DB 693 KGTVHARILKPSDLRKIKPLGSGVGTGSKFWIPEGETYKIPVALIKTIQDSSGRQTF 752
QY 772 EILDEAYMAGVSGSYVRLIGICTSTVOLVQTMVYGCILHVRNKRGLSODILNW 831
DB 753 EITDHLMSGSLDHYIVRLIGICGTQLQVTLSSHGLSHLEIRKHSLSLPQRLNW 812
QY 832 CMQIAKMSYLEDEVLYARDLAARNVLYKSPNHKITDFGLARLDDDETEYHADGGKVP 891
DB 813 CVQIAKGMVLYBEHRVVKHLLAARNILKNDYQVQISDYVADLLYDDKKVYSEKTP 872
QY 892 IKMMALESILRRRPHQSDVWSYGVTTWELMTFGAKPYDGIIPAREIDOLEKGRLEPP 951
DB 873 IKMMALESILRRRPHQSDVWSYGVTTWELMTFGAKPYDGIIPAREIDOLEKGRLEPP 951
QY 952 ICTIDVYIMVYCMWIDSECRPRELTVSEPSRMAPORPVVIONDLPASPDLSTFY 1011
DB 933 ICTIDVYIMVYCMWIDSECRPRELTVSEPSRMAPORPVVIONDLPASPDLSTFY 1011
QY 1012 RSLLEDDMDGLVDAEYLVPOQGFCCPAPAGAGVHHRHRSSTRSGGDLTLGLEP 1071
DB 981 ---EDSGMGEFL-----RGSER---GILEADLEE 1004
QY 1072 SEEEAPRSPLAPSEGASDVFDGLGNG--AAKGLQSLPHDSPLQ-----RY 1118
DB 1005 DEEB-----GLGDRFAPSLPSPBSWSTSPQINSYMMTQLAR 1043
QY 1119 SEDPTVPLPSETDGVAVATLCSPOP-EYVNO-----PVRPQPPSPREGPL 1163
DB 1044 D-----FAVSGGHHIYGLPMSPSVDITRQLMYORSLSSVRLPDRSAFRSSSEAE 1097
QY 1164 --PAARPAATLERKKTISPGKNGVMDVFAFGAVENPELITFOGGAALQPHPPAFSP 1221
DB 1098 CEDGAQCAGIFRVR-----FGSERGN-----FOGG----- 1122
QY 1222 AFDNLVYWDODPPERGAPSTFKGPTAENPE 1253
DB 1123 -----QQRKLSTASSPSSFKTMADEDE 1146

```

RESULT 7
Q9BIH9
ID Q9BIH9 PRELIMINARY; PRT, 1433 AA.
AC Q9BIH9;

```

DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SU4.
RA Lysate G.U.
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor."
RT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301655; CAC35008.1; -
DR HSRP; P11362; 1FKK.
DR InterPro; IPR000345; Cytochrome_b.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Kinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_kinase; 1.
DR SMART; SM00261; Fu; 7.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON_TER.
SQ
SEQUENCE 1433 AA; 159585 MW; E3D9D8896772AF07 CRC64;
Query Match 29.6%; Score 2030.5; DB 5; Length 1433;
Best Local Similarity 32.3%; Pred. No. 2.8e-144;
Matches 470; Conservative 199; Mismatches 393; Indels 391; Gaps 36
QY 26 CTGDMKRLTASETHLDMLRHYQCGVYQNLLETTPYTNASLSFLDDIQEVGYVL 85
DB 1 CIGTNGRMSVAPNREHYKMLRDRYTNCTYVDGDLITWTIONITDINFLQIREVGYVL 60
QY 86 IAHNVGRVPLQRIYRGTLF-----EDNVALAVLDNDGDPINNTTPVYGASPGGLRE 140
DB 61 ISLYDLPOVILLPRLQITIRGRITFKLNMKEEAYGLFV-----SFSHMTL 104
QY 141 QLRSTIELKGLVLIQRNPOLCYODTILMKDI-FHKNQALATLIDNRSRACHPSPMC 199
DB 105 ELPALRDLIGSVGFNNYMLCHVKSINMEIILAPQTSQYTFNFSPRVPCHPSC 164
QY 200 KGSRCWGSSESSDQSLRRTVAGGCA--RCKGRLPDCHECCAGAGCTGKSHDCLACH 257
DB 165 EVG-CWGSAGNCRFSKLNCSPOCSGCGPGRPCCHLPAGCGTGTOGSDCLACKN 223
QY 258 FNHSGICELHCPALVTYNTDIFESMPNDEGRYTFGASCYACAPYNYLSTDVSGCTLVCP 317
DB 224 FYDDGVCKQCECPMQIYNPTNYFMEPRPDKXVAGATCVAKC-BHLKMGNGACVAKCPR 282
QY 318 HNGEYTAEDGTQREKSKPCARCYGLGNFNNFTVSWLRLVRSASHLEFPAGCKKIF 377
DB 283 GKMPQNSE-----CVPKGVCPKTCPEGI-----VHSDNGYKDCDTIIE 323
QY 378 GSLAFIPESFPGDPAASNT-----APLQPEQLOVEFTLEIGLYISAWDLSPLDS 429
DB 324 GSLRELDSFPGPOGVYTNFSFGFRYKIKIPDRLEVPSTYKELTGFTINIOAHNPFTTLN 383

```

```

QY 430 VFONLOYRGRILBNQAY-STLTQGLGISWLGSRISREISGLALINHTHLCFVHTVPM 488
DB 384 YFRNLEIVYVGRQLKENIPASVYIVKXSLKSLKSLKRVNVSIVLENSDDCFVEDIOM 443
QY 489 DQLEFRNPHQALLHTANPEDECEVCEGLACHQLCARHGCHWPGRTQCVNCSQFLRQCECE 548
DB 444 SEIKSSSDHEVMVQKNNNATECHEGMECESEOCSSKAGCWAKGEQCLEKNVYKXKCID 503
QY 549 ECRVLOGLPREY-VNABHCLPCHPECCOPONGSVTCFSPEDQCVACAHYDPPFCVAPCP 607
DB 504 SCK--SLPRLYSDVKTCGDCHQECQ-----FCYSPNEDNGSCMANYKDGFCVACBEP 555
QY 608 -----SGVXPDLSTYMTWKFPD-- 624
DB 556 TTKANNGTCINCHKTVCGRGPRDTIAPDGISCDKAIIGSDAKIERCLMKDSCPDGY 615
QY 625 -----EEG----- 627
DB 616 YSDVVLQEEGPKQLSGKAVCRKCHPRCKCTGYGFHEFCQECTGYKKGEQCEDECPD 675
QY 628 -----ACOPCPINCT-----HSCVDL-----DD-----KCGPAAQ-- 652
DB 676 FYANEEETRICLPCHQECGCHGLDGDHHECRNLKLFEGDPYDNATPTTCVSNCPASHPYK 735
QY 653 -----RASPLSTYSAVVGLLVVVLGVVFGI-----LIKRRQK 687
DB 736 RPEOAGKIGPYCASDMSQSGRLIEPQTQVKIWMGSVALILCVVFGIAFVLPFRHKNK 795
QY 688 IKKYTMRRLQETELVEPLTPSGAMPNOAMRILKETELRKVKYKLSGAGATGYKGIWIP 747
DB 796 KDAVKYTMALAGCESEPLRPSNNGPNITKRIITIKENIRGVLGNGAGORVFKGVWMP 855
QY 748 DGENVKIPVAIKVLENTSPRANKELDEAYVAGVSPYSLGICLSTVQLVQLM 807
DB 856 ESESVKIPVAIKVLENTSPRANKELDEAYVAGVSPYSLGICLSTVQLVQLM 915
QY 808 FYGCLLDHYRENGRLGSDLLNMCQIAKGSYEDVRLVHRDLAARNLVKSPNHVKI 867
DB 916 PLGCLLDHYRENGRLGSDLLNMCQIAKGSYEDVRLVHRDLAARNLVKSPNHVKI 975
QY 868 TDFGLARLLIDETERYHADGKRVPIKMALESILRRFTQSDVWSYGVTVWELMTGAK 927
DB 976 TVFGLAKLLDPDSDEYRAAGKVPKIKMLALBICIRHVRFTSKSDWAFFGITWELITGAR 1035
QY 928 PYDGIAPARIPDLLEKGERLPPPCITIDVYMWKCMIMISEGRPRRELVESEFMMAR 987
DB 1036 PYENVPKADVPPELIEGHKLPPPDICSDVYICILSCWVLDDADAPYFKQLAETFAEKAR 1095
QY 988 DPQRFVYIQNEDLGPASPLDSTFYRSLLEDDMDGLV----- 1024
DB 1096 DPGRYIMI-----PDGKFMRLPSYTNQDEKDLITFLAFVMAAAAAAGASNV 1146
QY 1025 -----DABEYLVPOOGFCFPPRPAQAGMTHRRSSSTSGGDDLTLGLEPSEEEFMR 1078
DB 1147 VPETIAETDEYLOPKRPSIMLPSPA-----VEPS-DEMPK 1182
QY 1079 S-----PLAP--SEGASDVFDGDLGKAGAKQLSLPTHDPSPLOQYSEDPVPLPSE 1129
DB 1183 SLRYCKDPLKPDDETGHGKEV-----GVGGIR-----LNLPLD 1216
QY 1130 TDGYVAPLTCSPQREYVNPQVPPQPPSPREGRPLPAARPAGATLERAKTISPKGNKVYKD 1189
DB 1217 EDDYLMF-TCOSQ-----NOS-----TPG-----YMD 1237
QY 1190 VFAFGAVENREYL-----TPQGAAPQHPHPAPAFADNLVYWDQDPPERGAP 1240
DB 1238 LIGVPASVDNPEILMSTQAIAGLAGSNG--PHTP-----PF 1274
QY 1241 STEKGTPTAENPE 1253
DB 1275 NTENGMPTHQHSQ 1287

```

```

RESULT 8
Q9UK79 ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=9415951; PubMed=10485918;
RA Doherty J.K., Clinton G.M., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor."
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).
RN (2)
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF177761; AAD56009.2;
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; F6CIBB347E2D030C CRC64;

Query Match 27.3%; Score 1872.5; DB 4; Length 419;
Best Local Similarity 86.5%; Pred. No. 4e-133;
Matches 354; Conservative 6; Mismatches 27; Indels 13; Gaps 1;

QY 1 METALCWMGLLALLPFGAASVCTGTDMKRLPASPETHDMLRLYQCCVVOGNL 60
DB 1 METALCWMGLLALLPFGAASVCTGTDMKRLPASPETHDMLRLYQCCVVOGNL 60
QY 61 ELTYPTNASSFLQDIOEVGVYLAHNOVROYPLRLRVGTOLPEFNVALAVDNG 121
DB 61 ELTYPTNASSFLQDIOEVGVYLAHNOVROYPLRLRVGTOLPEFNVALAVDNG 121
QY 121 DPLNNTTPTVGTASPGGLRELOLRSTELIKGVLIORNPOLCYODTILMKDIFPKXNQLA 181
DB 121 DPLNNTTPTVGTASPGGLRELOLRSTELIKGVLIORNPOLCYODTILMKDIFPKXNQLA 181
QY 181 LFLIDTNRSPACHPCSPYCKSGRCKGSESSDCCSLTRVYCGAGCARCKGPLPTCCHEQC 241
DB 181 LFLIDTNRSPACHPCSPYCKSGRCKGSESSDCCSLTRVYCGAGCARCKGPLPTCCHEQC 241
QY 241 AAGCTGPKKSDCLALHNHSGICELHPALVTVNTDFFESMPNPEGRTYFGASCYIACP 301
DB 241 AAGCTGPKKSDCLALHNHSGICELHPALVTVNTDFFESMPNPEGRTYFGASCYIACP 301
QY 301 YNYLSTDVSGCTLVCPPLHNOEYTAEDGTQRCCKSKPCARVYCYGLGFMNNTTVFWLRVP 361
DB 301 YNYLSTDVSGCTLVCPPLHNOEYTAEDGTQRCCKSKPCARVYCYGLGFMNNTTVFWLRVP 361
QY 361 KYASHLLEFFACCKKIFGSLAFLPSPGDDASNPALQ 400
DB 360 PGPAHFV-----LSFLRPSMDLVSAFYSPLAP 387

RESULT 9
Q8R2X1 ID Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.

```


OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC027080; AAH27080.1;
 KW Hypothetical protein.
 SQ SEQUENCE 367 AA; 40163 MW; 0BE03395FE101B0 CRC64;

Query Match 25.4%; Score 1739; DB 11; Length 367;
 Best Local Similarity 88.0%; Pred. No. 4.1e-123;
 Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 895 MALESILRRRFTHQSDVSYGVTWELMTFGAKPYDGIPIAREIPDLLEKGERLPPICT 954
 DB 1 MALESILRRRFTHQSDVSYGVTWELMTFGAKPYDGIPIAREIPDLLEKGERLPPICT 60
 QY 955 IDVYIMVYKCMWIDECRPPRELVSERSMARDPQRFVVIQNEDELGPASPLDSTYRSL 1014
 DB 61 IDVYIMVYKCMWIDECRPPRELVSERSMARDPQRFVVIQNEDELGPASPLDSTYRSL 120
 QY 1015 LEDDMGDLVDAEEYLVPOQGFCDPPAPGAGWVHRHRSSTRSGADLTGLPESSE 1074
 DB 121 LEDDMGDLVDAEEYLVPOQGFCDPPAPGAGWVHRHRSSTRSGADLTGLPESSE 180
 QY 1075 EAPRSPLAPSGAGSDVFDGLGAKAGLQSLPTHDPSPLOQRYSEDPVLPSETDGYV 1134
 DB 181 EAPRSPLAPSGAGSDVFDGLGAKAGLQSLPTHDPSPLOQRYSEDPVLPSETDGYV 240
 QY 1135 APLTCSQPEYVNOVDVPOPPSPREGPLPARAPAGTLERAKTLSPGKGVKQVAFAG 1194
 DB 241 APLTCSQPEYVNOVDVPOPPSPREGPLPARAPAGTLERAKTLSPGKGVKQVAFAG 300
 QY 1195 GAVENPEYLTPOGGAAPQPHPPAFSPAFDNLVYMDODPEPBGAPSTFGTGTAEINPEX 1254
 DB 301 GAVENPEYLTPOGGAAPQPHPPAFSPAFDNLVYMDODPEPBGAPSTFGTGTAEINPEX 360
 QY 1255 LGLDVPV 1261
 DB 361 LGLDVPV 367

RESULT 10

Q86712 PRELIMINARY; PRT; 729 AA.
 AC Q86712;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Polypeptide.
 GN POLYPEPTIDE.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
 OX NCBI_TaxID=11950;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; PubMed=8152791;
 RA Venstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
 RA Johnson A., Beug H.;
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities."
 RL Oncogene 9:1307-1320(1994).
 DR EMBL; S69372; AAC60725.1;
 DR HSSP; P03322; 1A6S
 DR InterPro; IPR000719; Euk_kinase.
 DR InterPro; IPR004028; Retro_M.
 DR InterPro; IPR001245; Tyr_kinase.
 DR Pfam; PF00069; Kinase; 1.
 DR Pfam; PF02813; Retro_M; 1.

DR ProDom; PD000001; Euk_kinase; 1.
 DR SMART; SM00219; TyKc; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.
 DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
 KW ATP-binding; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EF81D63 CRC64;

Query Match 25.1%; Score 1720; DB 15; Length 729;
 Best Local Similarity 54.8%; Pred. No. 3.1e-121;
 Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 1;

QY 575 PONGSVTFGEADQCAVCAHYKDPPEVCVACRSGVCFDUSWPIKFPDEEGACQCPPI 63
 DB 141 PEETATPRTGP--DHCKCAHFIDGPHCVACAPAGVIGENDTL-VMKYADANVQCQLCHP 19
 QY 635 NCTHSCVLDLQKGPAGBAPLSTISAVV-GILLVAVGVGVGLIKRQOKIRKRYTM 69
 DB 198 NCTHSCVLDLQKGPAGBAPLSTISAVV-GILLVAVGVGVGLIKRQOKIRKRYTM 25
 QY 694 RRLQETELVPLTPSGAMPNQAQMRILKETELRKVVLGSGAFVYKGIWIPDGENVK 75
 DB 254 RRLQETELVPLTPSGAMPNQAQMRILKETELRKVVLGSGAFVYKGIWIPDGENVK 31
 QY 754 IPVALKUTRENTSRKANKEILDEAYVWAGSPYVSLGICLTSTVQLTQLMFYGCL 81
 DB 314 IPVALKUTRENTSRKANKEILDEAYVWAGSPYVSLGICLTSTVQLTQLMFYGCL 37
 QY 814 DHVENRGRSLQDILLWCMQIAGKMSYLEDVRLVHRDLAARVLYKS PNVKRTIDFGLA 87
 DB 374 DYIRHKONISQVLLWCMQIAGKMSYLEDVRLVHRDLAARVLYKS PNVKRTIDFGLA 43
 QY 874 RLUDIDEYHADGGKVPDKMALESLRFRFTHQSDVSYGVTWELMTFGAKPYDGI 93
 DB 434 RLUDIDEYHADGGKVPDKMALESLRFRFTHQSDVSYGVTWELMTFGAKPYDGI 49
 QY 934 AREIPDLLEKGERLPPICTIDVYIMVYKCMWIDECRPPRELVSERSMARDPQRFV 99
 DB 494 AREIPDLLEKGERLPPICTIDVYIMVYKCMWIDECRPPRELVSERSMARDPQRFV 55
 QY 994 VIQ-NEDELGPASPLDSTYRSLLEDMDGDLVDAEEYLVPOQGFCDPPAPGAGWVHNR 105
 DB 554 VIQ-NEDELGPASPLDSTYRSLLEDMDGDLVDAEEYLVPOQGFCDPPAPGAGWVHNR 59
 QY 1053 HRSSSTRSGGDLTLGLPESSEAPRSPL-----APSEGAGSDVFDGLGAKAGLQSL 110
 DB 599 NSPST-----SRTPLLSSLSATSNNSATNCID-----RNGQGH 631
 QY 1108 PTHDPSPLOQRYSEDPVLPSET--DGTVAPLTCSPQPEYVNOVDVPOPPSPREGPLPA 116
 DB 632 PTHDPSPLOQRYSEDPVLPSET--DGTVAPLTCSPQPEYVNOVDVPOPPSPREGPLPA 675
 QY 1166 ARPAGATLERAKTLSPGKGVKQVAF-----AFGAVENPEYL 1203
 DB 676 -----TAMVQNCIYNNISLTAISKLPMDSRYNQSHSTAVDNEYL 715

RESULT 11

Q86714 PRELIMINARY; PRT; 567 AA.
 AC Q86714;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE V-erbB protein (Fragment).
 GN V-ERBB.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
 OX NCBI_TaxID=11950;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; PubMed=8152791;
 RA Venstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,

RA Johnson A., Beng H.;
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities.";
 RL Oncogene 9:1307-1320(1994).
 DR EMBL: S69372; AAC60727.1; .
 DR HSSP: P1362; IFGC.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR001245; Tyr_kinase.
 DR Pfam: PF00069; kinase; 1.
 DR PRINTS: PRO0109; TYRKINASE.
 DR ProDom: PD000001; Euk_kinase; 1.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE: PS50011; PROTEIN KINASE DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KW Tyrosine-protein kinase.
 FT NON_TER
 SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7AD725E1 CRC64;
 Query Match 25.1%; Score 1718; DB 15; Length 567;
 Best Local Similarity 55.4%; Pred. No. 3e-121;
 Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;
 QY 584 GPEADQCAAHAYKPPPCVACRPSGVKPDLSYPIWKFPDEBACQPCPINCTHSCVDL 643
 DB 1 GP--DHCKCAHFIDGPHCVACRPAVAGENDTL-VWKYADANAVCCLCHPNCRCGKGP 57
 QY 644 DEKGPAGBRAPLRSIVSAVY-GILVWVGVGGLIKERQCKIRKYMRLLOETEL 702
 DB 58 GLEGP---NSKTSIAAGVGVGLICLVVGLIGLYLR--HIVKRLRLQREL 113
 QY 703 VEPLTPSGAMPNQAQMRILKETELKRVYLCGAGFYVYKGIWIPDENYKIPVALKVL 762
 DB 114 VEPLTPSGAPVQAALRIKETEFKRVYLCGAGFYVYKGLWPEGEKVIPIVALIELR 173
 QY 763 EHTSPKANKELIDEYVNAVGVSPVSHLIGICTSTVQVLTQMPGCLLDHRENGR 822
 DB 174 EATSPKANKELIDEYVNAVGVSPVSHLIGICTSTVQVLTQMPGCLLDYRHKCN 233
 QY 823 LGSODLLNMCQIAKMGSYLDEVRLVHRDLAARNLVKSPNHVXITDFGLARLDIDETE 882
 DB 234 IGSQYLNCVQIAQMGVYLERLVHRDLAARNLVXTPGHVXITDFGLAKLGADEKE 293
 QY 883 YHADGKVPYKMALESILRRFTHQSDVWSYGVTVWELMTFGAKYPDGIPIAREIPLE 942
 DB 294 YHAEKGKVPYKMALESILRRFTHQSDVWSYGVTVWELMTFGSKPYDGIPIAREISVLE 353
 QY 943 KGERLPQPICTIDYVMVWCMIDSECPREFRELSEFSRMARDPQRFVVIQ--NEDLG 1001
 DB 354 KGERLPQPICTIDYVMVWCMIDADSPREFRLIAFEKMARDPRIYVIGCDERH 413
 QY 1002 PASPLDSTFYRSLDEDDMGDIYDAEYLVVPOQGFCDPAFGAGVYHHRSSSTRSG 1061
 DB 414 LPSPTDSKFRYRLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
 QY 1062 GGDULTGLEPSEEAAPRPL-----APEEGAGSDVDFDGLGAKGAGLQSLPTDPSPLQ 1116
 DB 455 -----SRTPLLSLATSNNAKTNID-----RNGQGHVAVEDSFVQ 491
 QY 1117 RYSEDPVLPSET--DGYVADLTCSPOPEYVNPQVPPSPREGPLPAAPAGATLE 1174
 DB 492 RYSDPTGNFLEESIDDFL-----PAPEYVNO--LMPKKPS----- 526
 QY 1175 RAKTLSPGKGVVXDF-----AFGAVENPEYL 1203
 DB 527 ----TAMVONQIYNNISLTAISKLPMSDRYQNSHSTAVDNEEYL 566
 RESULT 12
 Q8WYVO PRELIMINARY; PRT; 412 AA.
 AC Q8WYVO;

DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 44.7 kDa protein.
 GN P13659.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
 RA Wan D.F., Gu J.R.;
 RT "Novel human cDNA clones with function of inhibiting cancer cell
 RT growth.";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF318349; AAJ55956.1; .
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR001245; Tyr_kinase.
 DR InterPro: IPR004019; YLP motif.
 DR Pfam: PF00069; kinase; 1.
 DR Pfam: PF02757; YLP; 2.
 DR ProDom: PD000001; Euk_kinase; 1.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;
 Query Match 24.8%; Score 1697.5; DB 4; Length 412;
 Best Local Similarity 80.5%; Pred. No. 6.6e-120;
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4
 QY 895 MALBSILRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPICT 954
 DB 1 MALBSILRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPICT 60
 QY 955 IDYVMVWCMIDSECPREFRELSEFSRMARDPQRFVVIQ--NEDLG 1001
 DB 61 IDYVMVWCMIDSECPREFRELSEFSRMARDPQRFVVIQ--NEDLG 120
 QY 1015 LEDDDMGDLVAEYLVVPOQGFCDPAFGAGVYHHRSSSTRSGGGLTGLEPSEE 107
 DB 121 LEDDDMGDLVAEYLVVPOQGFCDPAFGAGVYHHRSSSTRSGGGLTGLEPSEE 160
 QY 1075 EAPRSPLAPSGAGSDVDFDGLGAKGAGLQSLPTDPSPLQRYSEDPVLPSETDGYV 113
 DB 181 EAPRSPLAPSGAGSDVDFDGLGAKGAGLQSLPTDPSPLQRYSEDPVLPSETDGYV 240
 QY 1135 APLTCSPOPEYVNPQVPPSPREGPLPAAPAGATLEBAKTLSPGKGVVXDFVAFG 115
 DB 241 APLTCSPOPEYVNPQVPPSPREGPLPAAPAGATLEBAKTLSPGKGVVXDFVAFG 300
 QY 1195 GAVENPEYLTQGAAPQF-----HPPPA---PSPADNL 122
 DB 301 GAVENPEYLTQGAALSPFTLLPSAQPTSTINGRTHQSGGLHAPSKGHLRQRTQST 360
 QY 1227 YVWD-QDPPER-----GAPPSFKTPTPAEN 1251
 DB 361 WMTQCBQBEQVRRSPDVSSREGLTSAQIKRMEGPPPTTSRGTCHARN 410
 RESULT 13
 Q64895 PRELIMINARY; PRT; 962 AA.
 AC Q64895;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Gag, v-erb-A, v-erb-B protein.
 GN GAG-V-ERB-A-V-ERB-B
 OS Avian erythroblastosis virus.

OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
 NX NCBI_TaxID=11861;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9020603; PubMed=1969616;
 RA Brustkin A., Jackson J., Bishop J.M., McCauley D.J., Schatzman R.C.;
 RT "Six amino acids from the retroviral gene gag greatly enhance the
 RT transforming potential of the oncogene v-erb-B-";
 RL Oncogene 5:15-24(1990).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 DR EMBL: X52209; CAA36459.1; -.
 DR HSSP: P10828; 2NLL.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR000536; Hormone_rec_11g.
 DR InterPro: IPR001723; Steroid_receptor.
 DR InterPro: IPR001245; Tyr_kinase.
 DR InterPro: IPR001628; Znf_C4steroid.
 DR Pfam: PR00069; hormone_rec; 1.
 DR Pfam: PR00105; zt-C4; 1.
 DR PRINTS: PR00398; STRDHOMONER.
 DR PRINTS: PR00047; STROIDFINGER.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR ProDom: PD000035; Znf_C4steroid; 1.
 DR SMART: SM00430; HOL1_1.
 DR SMART: SM00219; Tyrc; 1.
 DR SMART: SM00399; Znf_C4; 1.
 DR PROSITE: PS00031; NUCLEAR RECEPTOR; 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE: PS00011; PROTEIN KINASE DOM; 1.
 DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.
 KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
 KW Transcription regulation; Transferase; Tyrosine-protein kinase;
 KW Zinc-finger.
 SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E4E95CE CRC64;
 Query Match 24.2%; Score 1660; DB 15; Length 962;
 Best Local Similarity 51.7%; Pred. No. 1.6e-116;
 Matches 358; Conservative 73; Mismatches 142; Indels 119; Gaps 18;
 QY 547 VEECRVQLGDPRE-YVNAR-HCLP-----CHPEQC 574
 DB 354 IEKCESTYLLAFEHYINRKINIDHFWSKILMKVADLMTGAYHASFHMKVECTELS 413
 QY 575 PONGSVTCFGEADQCVACAHYKDPFCVACPSGVKPDLSYPIKFPDEGACQPCPI 634
 DB 414 PQE-----VGP--DHCKKCAHFIDGPHCVACPCPGVLGENDTL-VMKYADANAVCOLCHP 465
 QY 635 NCHSCVDLDDKGPAPQORASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKYM 693
 DB 466 NCTRGCKGPGLEGCP---NGSKTPSIAGVVGGLLVVGLGLGLYLRR-HIVYKRTL 521
 QY 694 RLLQETELVEPLTPSGAMPNQAOMRLIKETELRKAVYLCSGAFGVYVGIWIDGENYK 753
 DB 522 RRLQERLVEPLTPSGEAPNQAHLRIKETEFKRYVGLFGAFGVYVKGWLPBEGEKT 581
 QY 754 IPVAIKVLRNENTSPKAKEILDEAYVAVGSPVYSRLIGICTSTVQVLTOLMPYGCIL 813
 DB 582 IPVAIKELRENTSKAKKEILDEAYVAVASVDNPHVCRLLGICLTSTVQVLTOLMPYGCIL 641
 QY 814 DHVKNRGLASQQLLWMCQIAKMSYLEDVRLVHRDLAARNVLSFNHYKITDFGLA 873
 DB 642 DYIREHKDNIQSQYLLWMCVQIAKMYLLEERHMYHDLAARNVLTQPHYKITDFGLA 701
 QY 874 RLDDIDETEVHADGKVPKIMMALESTLRRTFHOSDWSYGTVWELMTFGAKPYDGI 933
 DB 702 KQLGADKEHYHAEKGKPIKMALESTLHRIYHOSDWSYGTVWELMTFGSKPYDGI 761
 QY 934 AREIPDLLEKGERLPPICTIDYIMIMVYKMWIDSECRPFRELVSEPSRYARDPQREV 993
 DB 762 ASEISVLEKGERLPPICTIDYIMIMVYKMWISGADSPKRFELIAFSPKMAADPRL 821

QY 994 VIQ-NEDICPASPFLDSTFYRSLLEDDMDGLVDABEYLVPOQGFCCPPAPAGAMVHR 10
 DB 822 VIQGERMHLPSPTMSKRYRTLMEDMEDVDADEYLVPOGFF----- 86
 QY 1053 HRSSTREGGGLTLGLPSEEARSPLEASGAGSDVDFDGLMGAKQLGSLPTHDP 11
 DB 867 -NSPFT-----SRTPLLSLSATSN-----NSATKCIDRNGCH-- 89
 QY 1113 SPLQVSDPTVPLPSFTDGVAPLTCSPQEVYNOQDVRQPPSPREGLPAPAPGAT 11
 DB 899 -----FVREDGFL-----PAEEVYNQ--LMPKKESTAMVONQVNYISLT 93
 QY 1173 -LERAKTLSPGNKGVKDVFAFGAVENPEYL 1203
 DB 937 AISKLPMSRYQN-----SHSTAVNPEYL 961
 RESULT 14
 Q85468
 ID Q85468 PRELIMINARY; PRT; 545 AA.
 AC Q85468;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE Avian erythroblastosis virus (Ts34) v-erbB gene.
 OS Avian erythroblastosis virus.
 OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
 NX NCBI_TaxID=11861;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88217326; PubMed=2897102;
 RA Scottling P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
 RT "Common site of mutation in the erbB gene of avian erythroblastosis
 RT virus mutants that are temperature sensitive for transformation.";
 RL Oncogene Res. 1:265-276(1987).
 DR EMBL: X06943; CAA30024.1; -.
 DR HSSP: P11362; 1R6K.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR001245; Tyr_kinase.
 DR Pfam: PF00069; pkinase; 1.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE: PS00011; PROTEIN KINASE DOM; 1.
 DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.
 KW ATP-binding; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 545 AA; 60899 MW; 140DCE8C0A0F8AF4 CRC64;
 Query Match 24.0%; Score 1645; DB 15; Length 545;
 Best Local Similarity 54.9%; Pred. No. 9.3e-116;
 Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15
 QY 584 GREADQCVACAHYKDPFCVACPSGVKPDLSYPIKFPDEGACQPCPICTHSCVDL 643
 DB 1 GP--DHCKKCAHFIDGPHCVACPCPGVLGENDTL-VMKYADANAVCOLCHPNTRGCKP 57
 QY 644 DDKCPAPQORASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKYMRLQETEL 702
 DB 58 GLEGC---NGSKTPSIAGVVGGLLVVGLGLGLYLRR-HIVYKRTLRLQEREL 113
 QY 703 VEPITPSGAMPNQAOMRLIKETELRKAVYLSGAFGVYVGIWIDGENYKIPVAIKVLR 762
 DB 114 VEPITPSGEPNQAHLRIKETEFKRYVGLFGAFGVYVKGWLPBEGEVTTIPVAIKELR 173
 QY 763 ENTSPPKANKEILDEAYVAVGSPVYSRLIGICTSTVQVLTOLMPYGCILDHVENRGR 822
 DB 174 EATSPKANKEILDEAYVAVASVDNPHVCRLLGICLTSTVQVLTOLMPYGCILDYIREHKON 233
 QY 823 LGSQDLTNMCQIAKMSYLEDVRLVHRDLAARNVLSFNHYKITDFGLALDDIDE 862
 DB 234 IGSQYLLNMCVQIAKMYLLEERHMYHDLAARNVLTQPHYKITDFGLAQGLADEKE 293

```

QY 883 YHADGKVPKMMALSLRRRTHSDVMSYGVTVWELMTGAKPYDGPAREPDLLE 942
D 294 YHAEKGKVPKMMALSLRRRTHSDVMSYGVTVWELMTGAKPYDGPAREPDLLE 942
QY 943 KGERLPQPICTIDVYMWKCMMDSECRPRELVEFSFMSAADPQRFVVIQ-NEDLG 1001
D 354 KGERLPQPICTIDVYMWKCMMDSECRPRELVEFSFMSAADPQRFVVIQ-NEDLG 1001
QY 1002 PASPLDSTFYSRLLEDDMDGDLVDAAEYVVPQGFECPPDPAAGAGVHHHRSSTTSG 1061
D 414 LPSPTDSKFRITLMEEDMEDIDVDAEYVLPQGF-----NSPT--- 454
QY 1062 GGDLLTGLBPESEABRSP-----APSEAGSDVDFDGLGMAKQGLSPTHDPEFLQ 1116
D 455 -----SRTPLSLSLATSNSATNCIDRNG-----H----- 481
QY 1117 RYSEDPVPLPSTIDGVAPLTCSPQPEVYNQPDVPRQPSRSEGPLAARPAQT-LEK 1175
D 482 -----PYREDEPL-----PAPEVYNQ-LMPKKPSTAMVQIQIYVLSITLAIK 523
QY 1176 AKTLSPQKNGVVDVPAFGAGAVENPEYL 1203
D 524 LPMDSRYQN-----SHSTAVDNPPEYL 544

RESULT 15
Q9WVF5 PRELIMINARY; PRT; 655 AA.
ID Q9WVF5;
AC Q9WVF5;
DT 01-NOV-1999 (Tremblre). 12, Created)
DT 01-NOV-1999 (Tremblre). 12, Last sequence update)
DT 01-JUN-2002 (Tremblre). 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).
GN EGFR.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11)
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielson A.J., Scheil C.,
RA Lampard A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maibie N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors."
RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
RN 12)
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTA;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearshall R.S., Green P.J., Yee D., Lampard A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maibie N.J.;
RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egfr transcripts encoding truncated receptor isoforms."
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
RN 13)
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=1217851;
RA Kawal U., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Akakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batcatov S., Casavant T.,
RA Fetschmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schraml L.M., Stabli F., Suzuki R., Tomita M., Wagner L., Mashio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Hoffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

```

```

PA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
PA Gustincich S., Hill D., Hofmann M., Hume D.A., Kantiya M., Lee N.H.,
PA Lyons P., Marchionni L., Mashima U., Mazzarelli U., Mombereis P.,
PA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
PA Sasaki H., Sato K., Schenbach C., Seta T., Shibata Y., Storch K.F.,
PA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
PA Wyshaw-Borcia A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AF124513; AAD44149.1; -.
DR EMBL; AF275366; AAG28047.1; -.
DR EMBL; AF275364; AAG28047.1; JOINED.
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AK004944; BAB23688.1; -.
DR EMBL; AK004883; BAB23641.1; -.
DR EMBL; AK004911; BAB23662.1; -.
DR MGI; 95294; Egfr.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; Fuf; 3.
KM RECEPTOR.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 22.1%; Score 1512.5; DB 11; Length 655;
Best Local Similarity 44.2%; Pred. No. 1.3e-105;
Matches 286; Conservative 101; Mismatches 227; Indels 33; Gaps 14

QY 11 LIALIPGAA--STQVCTGDMKRLPASEPETHLMRLHYQCCVYQGNLELTYPTN 68
D 14 LTFALCAAGALEEKVKVQCGTSNRITQGTEDHFLSLQRMYNNECEVLGLLEITYQVRN 73
QY 69 ASLSFQDIOEYGVYVLAHNOVROFORIRYRGQLCFEDYALAVLDGDLNNTTP 121
D 74 YDLSFKTIQEVAGVYVLAHNTVERIPLENQIIRGNALVNTALALISL----- 12
QY 129 VTGASPGQIRELQRLSLTEILKQVGLIGRNPQLCYQDTILMKDI-----FHRNQLALTLI 18
D 125 -YGNRTGRLRELPMNLQELIILGAVRFNNPILCNMDTIQRDIDVQVFNMSMDL--- 18
QY 185 DNNRSAGCHPCSPMKGSRMCWSESEDDQSILTRYCAGGA-RCKGPLPTCCHEGCAAG 24
D 181 -QSHPSCKCPSPSPGNSCWGCGEENQKLTIKTICAQCSHRGRGSPSCCHNQCAAG 23
QY 244 CTGPKNSDCLACLNHNSGICELCPALVTYNTDTPFESMPPEGRYTFGASCYACAPYNY 30
D 240 CTGPRESDCLVCQKQDQDQATCKDTPRLMLNPTTYQMDVNPBKYSFGATCVKCKRNY 29
QY 304 LSTDVGSCTVLPRLHNOEYTAEDGTQRCCEKSKPCARVYCYGLM--FNNFTVSWLRAVPK 36
D 300 VYTDHSGSCVRAQGPDPYEV-EDDGIRKCKKCDGCRKVCNIGIGGEFED-TLS----- 35
QY 362 VSASHLEBPAGCKKIFGSLAFIPESFDDPSANTAPLOPEQLGVETLLEITGLYISAW 42
D 351 INATNIKHYKXTALSGDLHLIPAFKDSSTRPRDLPRLELTKYKETSGLLQAM 41
QY 422 PSLDPLSVQNLQVIRIRILHNGAVSLTQGLISWLGRLSLRELSGLALHHNTHLC 481
D 411 PDMWTDLHAFENLLEIRIRRTXQHGQFSLAVGLNITSIGRLSKRISDGVYIIGNENLC 476
QY 482 FHTVPMQDLFENPQALHLHTANPREDCVGEGLAQHQLCARGCMKPGPTQCNCSQFL 541
D 471 YANTTNMKKLTFTPOQKTKIMNRAEKDCQAVNHVNCNPLGSECCWKPPEPDCVSCNVS 531
QY 542 RQGCVECEGRVYLGRLPREYVNAHRLCPCHPECPQNSVTCFGEADQCAVCAHYKDPF 601
D 531 RRECEVEKCNILGEPREPFVENSECTGHEPCLPQAMNITCTGKGPONCICQAHYIDGPH 590
QY 602 CVARPGSVKPDLSMPYMKRPDESGACQRPINICTSSCYDLDDXGC 648
D 591 CVKTPAGIMGENNTL-VWKYADANNVCHLCHANCTYCGAGPGLOGC 636

```

Tue Jul 22 12:40:30 2003

seq4-325-339-14.rpt

Page 1

Search completed: July 22, 2003, 09:01:06
Job time : 53.5887 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 38.1589 Seconds

(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-325-339-14

Sequence: 1 MELALCWMGULLALLPPGA.....TFKCTPAENPEYIGLDVPI 1261

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002:*

1: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6736	98.3	1255	21	AA92620 Human herexgulin 2
2	6736	98.3	1255	22	AAE12130 Human tyrosine kin
3	6736	98.3	1255	22	AAE60167 HER2 transgene pla
4	6736	98.3	1255	23	AAU74545 Human HER2 (Rb52)
5	6730	98.2	1255	17	AAU01111 HER-2/neu protein.
6	6730	98.2	1255	20	AAE24406 Human HER-2/neu on
7	6730	98.2	1255	21	AAE21198 Human HER-2/neu pr
8	6730	98.2	1255	21	AAE41780 Amino acid sequenc
9	6730	98.2	1255	22	AAE85458 Human HER-2/neu pr
10	6730	98.2	1255	22	AAE88267 HER2/neu amino aci

11	6730	98.2	1255	23	AAE24067 Human Her-2 protei
12	6730	98.2	1255	23	AAE20479 Human Her-2/neu pr
13	6730	98.2	1255	23	AAE51143 Human Her-2/neu on
14	6730	98.2	1255	23	AAU77114 Human Her-2/neu po
15	6670	97.6	1433	14	AAE39568 Sequence of c-erbB
16	6566	95.8	1223	23	AAU08923 Human breast cance
17	6413	93.6	1200	21	AAE21208 Human HER-2/neu pr
18	5935.5	86.6	1256	21	AAE21199 Rat HER-2/neu prot
19	5935.5	86.6	1256	23	AAE51144 Rat Her-2/neu onco
20	5910.5	86.3	1256	22	AAE21206 Mouse Her-2/neu pr
21	5910.5	86.3	1256	22	AAE28860 Amino acid sequenc
22	5910.5	86.3	1256	22	AAE51151 Mouse Her-2/neu on
23	4816	70.3	919	21	AAE21203 Human HER-2/neu fu
24	4816	70.3	919	23	AAE51148 Her-2/neu extracel
25	4060.5	59.3	920	23	AAE51152 Mouse Her-2/neu ex
26	4060.5	59.3	926	23	AAE51153 Mouse Her-2/neu ex
27	3700	54.0	712	21	AAE21204 Human HER-2/neu fu
28	3700	54.0	712	23	AAE51149 Her-2/neu extracel
29	3554	51.9	782	18	AAE19764 Her2-GM-CSF Immuno
30	3552	51.8	653	21	AAE21200 Extracellular HER-
31	3552	51.8	653	23	AAE51145 Human Her-2/neu on
32	3514	51.3	645	22	AAE60408 Human ErbB2 oncopr
33	3514	51.3	645	22	AAE61593 Human ErbB2 extrac
34	3449	50.3	951	21	AAE44993 DCsecFv-erbB2Fc fu
35	3346	48.8	624	11	AAE08222 Extracellular porc
36	3150	46.0	1210	21	AAE19259 Amino acid sequenc
37	3150	46.0	1210	21	AAE50616 Human EGF receptor
38	3150	46.0	1210	23	AAE23019 Human Her-1 protei
39	3150	46.0	1210	23	AAE50768 Human epidermal gr
40	3148	45.9	1210	22	AAE68450 Amino acid sequenc
41	3109	45.4	1210	22	AAE51768 Human epidermal gr
42	3084	45.0	583	23	AAE20463 Human protein for
43	3084	45.0	587	23	AAE20451 Human protein for
44	3083	45.0	589	23	AAE20484 Human protein for
45	3083	45.0	600	23	AAE20482 Human protein for

ALIGNMENTS

RESULT 1	AA92620	standard; Protein; 1255 AA.
ID	AA92620	
AC	AA92620;	
XX	10-AUG-2000	(First entry)
XX	Human herexgulin 2 (Her2).	
DE	Human herexgulin 2 (Her2).	
XX	Herexgulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;	
KW	self-protein; cancer; breast cancer; prostate cancer;	
KW	cell-associated peptide antigen; foreign epitope.	
XX	Homo sapiens.	
OS		
XX		
XX	Key	Location/Qualifiers
FT	Domain	1..173
FT		/label= N-terminal
FT		/note= "mature polypeptide"
FT	Region	5..25
FT		/label= insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	59..73
FT		/label= insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	103..117
FT		/label= insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	149..163
FT		/label= insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Domain	174..323

FT	Region	/label= Cysteine_rich_domain
FT	Region	210..224
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	250..264
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	324..483
FT	Region	/label= Ligand_binding_domain
FT	Region	325..339
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	369..383
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	465..479
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	484..623
FT	Region	/label= Cysteine_rich_domain
FT	Region	579..593
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	624..654
FT	Region	/label= Transmembrane_domain
FT	Region	632..652
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	653..667
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	655..1010
FT	Region	/label= Tyrosine_kinase_domain
FT	Region	661..675
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	710..730
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	1011..1235
FT	Region	/label= C-terminal_domain
PN	MO200020027-A2.	
PD	13-APR-2000.	
XX	05-OCT-1999;	99WO-DK00525.
XX	05-OCT-1998;	98DX-0001261.
XX	20-OCT-1998;	98US-0105011.
XX	(MEBT) M & E BIOTECH AS.	
XX	Seinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I,	
XX	Gautam A, Birx P, Karlsson G;	
XX	WPI; 2000-349917/30.	
XX	N-PSDB; AAA09455.	
XX	Inducing immune responses to weakly immunogenic, tumor associated	
XX	peptide antigens for the treatment of breast and prostate cancer	
XX	Claim 62; Page 193-198; 220pp; English.	
CC	This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of	
CC	Her2 can be used in the claimed method as an autovaccine to induce a CTL	
CC	response. Subdominant CTL epitopes, antibody binding regions and	
CC	cysteine residues involved in disulfide bonds are preserved in the	
CC	immunogenised forms. Regions suitable for the insertion of foreign	
CC	helper epitopes were identified (see features table). The method	

is used for inducing immune responses against weakly immunogenic cell-associated peptide antigens (PA) such as those associated with cancers (self-proteins), e.g. human prostate specific membrane antigen (PSM), heretulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b). The method comprises effecting simultaneous presentation by antigen producing cells (APCs) of the animals immune system of: (1) at least 1 CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1 B-cell group derived from the cell-associated PA; and (2) at least 1 first T helper cell group which is foreign to the animal. Analogues of human PSM, human Her2 and human/murine FGF8b comprising a substantial part of all known and predicted CTL and B-cell epitopes of the respective PA and including at least one foreign T helper epitope are also claimed. The method is used to treat prostate, prostate/breast or breast cancer when the PA is human PSM, FGF8b and Her2, respectively.

Sequence 1255 AA;

Query Match	98.3%	Score 6736;	DB 21;	Length 1255;
Best Local Similarity	98.6%	Pred. No. 0;		
Matches 1244; Conservative	7;	Mismatches	3;	Indels 8; Gaps 2

QY 1 MEI¹AL¹CRWG¹LL¹AL¹PG¹AST¹OV¹CTG¹TD¹KL¹RL¹PAS¹ETH¹LM¹RHL¹YQG¹CV¹QGN¹L 60

DB 1 MEI¹AL¹CRWG¹LL¹AL¹PG¹AST¹OV¹CTG¹TD¹KL¹RL¹PAS¹ETH¹LM¹RHL¹YQG¹CV¹QGN¹L 60

OY 61 ELTYLPTNASLSFLQDIEVGGYLIAHQVRQVPLQRIRIVRGQLFEDNYALAVLDNG 120
| | | | |
Db 61 ELYTLPTNASLSFLQDIEVGGYLIAHQVRQVPLQRIRIVRGQLFEDNYALAVLDNG 120
| | | | |

QY 121 DPLNNTTPTGTGAPGGLRELQRLTEIKGVLIQRNPQLCYQDTILMKDIFHKNNQLA 180
|||
Db 121 DPLNNTTPTGTGAPGGLRELQRLTEIKGVLIQRNPQLCYQDTILMKDIFHKNNQLA 180

QY 181 LTLIDTNSRACHCSPMCKGSRCKGESEDCSLTRTVAGGACRCKGBLPTDCHEQC 240

DB 181 LTLIDTNSRACHCSPMCKGSRCKGESEDCSLTRTVAGGACRCKGBLPTDCHEQC 240

[illegible]

```
QY      301 YNYLSTDVSSCTLVCPPLHNQEVTAAEDGTGRCCEKSKPCARVVCYGIGMFFNFTVSFWLR-V 359
        |||||
Db       301 YNYLSTDVSSCTLVCPPLHNOEVTAAEDGTGRCCEKSKPCARVVCYIGIMEH-----LREV 353
```

QY 360 PKVSASHLEEFACCKIFGSLAFLPESFGDPASNTAPLQPEQLVFETLEEITGYLYIS 419
:::~::~

Db 354 RAYTSANIQEAFACKKIFGSLAFLPESFGDPASNTAPLOPEQLVFETLEEITGYLYIS 413

OY **420** AWPDSLPIISVFONLQVTRGRILHNGAVSLTLOGLISMVLGRLRELSSGLALIHNTHT
|||
|||
Db **414** AMPDSSLPIISVFONLQVIRGRILHNGAVSLLTGIGISMVLGRLRELSSGLALIHNTHT **473**

QY 480 LCFVHTVPMDQLFRNP HQALLHTANRPEDECVGGLAQIQCARGHCWGPPTCCVNCSQ 539
Db 474 LCFVHTVPMDQLFRNP HQALLHTANRPEDECVGGLAQIQCARGHCWGPPTCCVNCSQ 533			

OY		540	FLLGQECVEECSRVIQGRLPREVYNARHCLPCHPECGPONGSVTTCGPBADQCVAACHKKDP	599
Ddb		534	FLLGQECEEBCRVIQGLRPRELYNARHCILCPHPCEOPONGSVTCCGPLADQCVAACHKDP	593

QY 600 PFCVACRCSGVKPDLSYMPIWKPEDEEGACQPCCINCTHSCVDLDDKCGPAEQRASPLTS 659
| | | | |
Db 594 PFCVACRCSGVKPDLSYMPIWKPEDEEGACQPCCINCTHSCVDLDDKCGPAEQRASPLTS 653

DQ 660 IYSAVGLLWVLGVFGLIKRROQKIRKYMRLLIQTBLVEPLTPSGAMNQAKNR 719
Db 664 IYSAVGILLWVLGVFGILIKRRQQKIRKYMRRLIQETELVEPLTPSGAMNQAKNR 713	

[illegible]

```

QY 780 MAGVSSPYVSRLLGICLTISTVQLVLTQMLPYGCLLDHVENKRGSLGQDLLNTMCQIAKGM 839
DB 774 MAGVSSPYVSRLLGICLTISTVQLVLTQMLPYGCLLDHVENKRGSLGQDLLNTMCQIAKGM 833
QY 840 SYLEDVRLVHRDLARNLVKSPPNHVKITDFGLARLLDIDTEYHAGGVPIKMALES 899
DB 834 SYLEDVRLVHRDLARNLVKSPPNHVKITDFGLARLLDIDTEYHAGGVPIKMALES 893
QY 900 ILRRRFTHQSDVMSYGVTVWELMTFGANPYDGIAPAREIPDLLEKGERLPQPICTIDVYM 959
DB 894 ILRRRFTHQSDVMSYGVTVWELMTFGANPYDGIAPAREIPDLLEKGERLPQPICTIDVYM 953
QY 960 IMVXCMIMDSECRPRFRELVSFBSMADDPQRFVYIQNEDIGPASPLDSTFYRLLEDDO 1019
DB 954 IMVXCMIMDSECRPRFRELVSFBSMADDPQRFVYIQNEDIGPASPLDSTFYRLLEDDO 1013
QY 1020 MGDLYDAEELVPOOGFCPPDPAPAGAGMVRHRSSTRSGGDLTLGLPSESEAPRS 1079
DB 1014 MGDLYDAEELVPOOGFCPPDPAPAGAGMVRHRSSTRSGGDLTLGLPSESEAPRS 1073
QY 1080 PLAPSEGAGSDVFDGDLGMAAKGIQSLPTHDPSPLOKRYSEDPTVPLPSETDGVAPLTC 1139
DB 1074 PLAPSEGAGSDVFDGDLGMAAKGIQSLPTHDPSPLOKRYSEDPTVPLPSETDGVAPLTC 1133
QY 1140 SPQPEYVNOPDVRPQPSRREGSPLPAARPAGATLEBAKTLSPGKNGVYKDVFAFGAVEN 1199
DB 1134 SPQPEYVNOPDVRPQPSRREGSPLPAARPAGATLEBAKTLSPGKNGVYKDVFAFGAVEN 1193
QY 1200 PEYLTPOGGAAPQHPPPAPFAPDNLVYWDQDPPERGAPSTFKGTPTAENPEYIGLDV 1259
DB 1194 PEYLTPOGGAAPQHPPPAPFAPDNLVYWDQDPPERGAPSTFKGTPTAENPEYIGLDV 1253
QY 1260 PY 1261
DB 1254 PY 1255

```

RESULT 2
AAE12130
ID AAE12130 standard; Protein, 1255 AA.

```

XX AAE12130;
AC 18-DEC-2001 (first entry)
DT 18-DEC-2001 (first entry)
XX
DE Human tyrosine kinase-type receptor, HER-2.
XX
KW Therapeutic compound; major histocompatibility complex; vaccine;
KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW antigen presenting cell; human; tyrosine kinase-type receptor.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 774..782
FT /note= "Antigenic epitope"
PD 20-SEP-2001.
XX
PF 16-MAR-2001; 2001WO-US40328.
XX
PR 16-MAR-2000; 2000US-0527487.
XX
PA (GENZ ) GENZYME CORP.
XX
PI Nicolette CA;
XX
XX WPI; 2001-616284/71.
XX DR N-PSDB; AAD19731.
XX

```

PT Novel synthetic therapeutic compound for inducing immune response and
PT for use in adoptive immunotherapy, has enhanced binding to major
PT histocompatibility molecules and enhanced immunoregulatory properties
PT

Claim 4; Page 63-67; 69pp; English.

CC The invention relates to synthetic therapeutic compounds (antigenic
CC peptides) with enhanced binding to major histocompatibility complex
CC (MHC) molecules and enhanced immunoregulatory properties relative
CC to their natural counterparts. Compounds of the invention are useful
CC for inducing an immune response in a subject and for use in adoptive
CC immunotherapy. They are useful as components of anti-cancer vaccines
CC and to expand immune effector cells that are specific for cancers
CC characterised by expression of the breast cancer antigen, HER-2.
CC Polynucleotides that encode peptides of the invention are useful as
CC hybridisation probes and as primers for the detection of genes of gene
CC transcripts that are expressed in antigen presenting cells (APCs), to
CC confirm transduction of polynucleotides into host cells. The present
CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC of the invention are designed based on the HER-2 antigenic peptide
CC (774-782).

XX Sequence 1255 AA;

Query Match 98.3%; Score 6736; DB 22; Length 1255;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1244; Conservative 7; Mismatches 3; Indels 8; Gaps 2;

```

QY 1 METALCRWGLILLALPRGASTVCTGTMKRLRPAPETHLDMRLHLYQGQVVGQNL 60
DB 1 METALCRWGLILLALPRGASTVCTGTMKRLRPAPETHLDMRLHLYQGQVVGQNL 60
QY 61 ELYLPTNASTLFDIDIEVQGYVLIANNQVQVPLQRLIVRGTOLEFENYALAVDNG 120
DB 61 ELYLPTNASTLFDIDIEVQGYVLIANNQVQVPLQRLIVRGTOLEFENYALAVDNG 120
QY 121 DPLNNTPTVTGASRQGLRELOLRSLTILKGVILIQNPOLCYODTILMKDIFPKNQLA 180
DB 121 DPLNNTPTVTGASRQGLRELOLRSLTILKGVILIQNPOLCYODTILMKDIFPKNQLA 180
QY 121 DPLNNTPTVTGASRQGLRELOLRSLTILKGVILIQNPOLCYODTILMKDIFPKNQLA 180
DB 121 DPLNNTPTVTGASRQGLRELOLRSLTILKGVILIQNPOLCYODTILMKDIFPKNQLA 180
QY 181 LTLIDTNSRACHCSPMKGSRMGSESEDCOSLRTVAGGAGARCKGLPTDCCHEOC 240
DB 181 LTLIDTNSRACHCSPMKGSRMGSESEDCOSLRTVAGGAGARCKGLPTDCCHEOC 240
QY 241 AAGCTGPRGSDCLALCFHNSGICEJLCPALVTYNTDTPESMENPEGRYTFGASCTYACP 300
DB 241 AAGCTGPRGSDCLALCFHNSGICEJLCPALVTYNTDTPESMENPEGRYTFGASCTYACP 300
QY 301 YNYLSTDVSGCTIWCPLHNOEVTAEDEGTQCEKSKRCACVCGLGMFNNFTVFWLR-V 359
DB 301 YNYLSTDVSGCTIWCPLHNOEVTAEDEGTQCEKSKRCACVCGLGMEH-----LREV 353
QY 360 PKVASHLSEFAGCKKIFGLAFLEBSFSDGDPASNTAPLOPELOVYFETLEETGLYIS 419
DB 360 PKVASHLSEFAGCKKIFGLAFLEBSFSDGDPASNTAPLOPELOVYFETLEETGLYIS 419
QY 419 PKVASHLSEFAGCKKIFGLAFLEBSFSDGDPASNTAPLOPELOVYFETLEETGLYIS 419
DB 419 PKVASHLSEFAGCKKIFGLAFLEBSFSDGDPASNTAPLOPELOVYFETLEETGLYIS 419
QY 420 AMPDLSPLSVFQWLQYIRGLIHNGAYSLTLQGLISWLGSLRELQSGALIHNNTH 479
DB 414 AMPDLSPLSVFQWLQYIRGLIHNGAYSLTLQGLISWLGSLRELQSGALIHNNTH 473
QY 480 LCFVHTVPMQDLFNNHQALLHTANREDECVGEGLAHQQLARGHCMGGRPOCNCGO 539
DB 474 LCFVHTVPMQDLFNNHQALLHTANREDECVGEGLAHQQLARGHCMGGRPOCNCGO 533
QY 540 FLRGQECVEGCVLQGLPREYVNAHCLPCHEPCOPQNGSVTCFGEADQCVACAHYKXP 599
DB 534 FLRGQECVEGCVLQGLPREYVNAHCLPCHEPCOPQNGSVTCFGEADQCVACAHYKXP 593
QY 600 PFCVARGSGVKKPDLSTVMPIMKFPDEGAGQPCPFINCTHSQVLDLXGCPABRASPILTS 659
DB 594 PFCVARGSGVKKPDLSTVMPIMKFPDEGAGQPCPFINCTHSQVLDLXGCPABRASPILTS 653

```

```

QY 660 IVSAVVGILLVVVLGVFGILLIKRQOKIRKXTYMRLLQETELVEPLTPSGAMPNOQMR 719
DB 654 IVSAVVGILLVVVLGVFGILLIKRQOKIRKXTYMRLLQETELVEPLTPSGAMPNOQMR 713
QY 720 ILKETELRKVKVLSGAFGTGKGIWIPGENVKIIPVAIKVLRNTSPKANKELIDEAYV 779
DB 714 ILKETELRKVKVLSGAFGTGKGIWIPGENVKIIPVAIKVLRNTSPKANKELIDEAYV 773
QY 780 MAGVSPVYSRLIGILTSTVOLVTLQMPYGCILHVRENDRGLSGOLIMCQOIAQM 839
DB 774 MAGVSPVYSRLIGILTSTVOLVTLQMPYGCILHVRENDRGLSGOLIMCQOIAQM 833
QY 840 SYLEDVLYHRDLAANVLYKSPNHVKITDFGLARLIDIDETEHADGKXPIKMMALES 899
DB 834 SYLEDVLYHRDLAANVLYKSPNHVKITDFGLARLIDIDETEHADGKXPIKMMALES 893
QY 900 ILRRRFTHSDVMSVGTVMELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 959
DB 894 ILRRRFTHSDVMSVGTVMELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 953
QY 960 IMVCMWIDSECRPRRELIVSEFSRWARDPQRFVYIIONEDLGPASPLDSTFYRSLLEDD 1019
DB 954 IMVCMWIDSECRPRRELIVSEFSRWARDPQRFVYIIONEDLGPASPLDSTFYRSLLEDD 1013
QY 1020 MGDVLDAAEYLVVQGFCDPDPAPGAGGVHHRHSSSTRSGGDLTGLPSEEBEAPRS 1079
DB 1014 MGDVLDAAEYLVVQGFCDPDPAPGAGGVHHRHSSSTRSGGDLTGLPSEEBEAPRS 1073
QY 1080 PLAPSEGAGSDVDGDLGMAAKGLQSLPTHDPSPQRYSEDPVPLPSETDGYVAPLTC 1139
DB 1074 PLAPSEGAGSDVDGDLGMAAKGLQSLPTHDPSPQRYSEDPVPLPSETDGYVAPLTC 1133
QY 1140 SPQPEYVNOQDVAPQPPSPREGPLPAAPAGATLEBAKTLSPGXGVYKDVAFAGAVEN 1199
DB 1134 SPQPEYVNOQDVAPQPPSPREGPLPAAPAGATLEBAKTLSPGXGVYKDVAFAGAVEN 1193
QY 1200 PEYLTPOGGAAPQHPHPAPFPAFDNLVYWDODPERGAPSTFKGTPTAENPEYLGIDV 1259
DB 1194 PEYLTPOGGAAPQHPHPAPFPAFDNLVYWDODPERGAPSTFKGTPTAENPEYLGIDV 1253
QY 1260 PV 1261
DB 1254 PV 1255

RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
XX
AC AAB60167;
XX
DT 03-APR-2001 (first entry)
XX
DE HRR2 transgene plasmid construct encoded protein.
XX
KM Human; HRR2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
XX antibody.
XX
OS Homo sapiens.
XX Synthetic.
XX
PN WC0200100244-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000MO-US17229.
XX
PR 25-JUN-1999; 99US-0141316.
XX 16-MAR-2000; 2000US-0189844.
XX
PA (GETH ) GENENTECH INC.
XX
PI Erickson S, Schwall R;

```

```

XX WP1: 2001-061962/07.
DR N-PSDB; AAF24297.
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid.
XX
PS Example 3; Fig 4; 92pp; English.
XX
CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HRR2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;
XX
Query Match 98.3%; Score 6736; DB 22; Length 1255;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1244; Conservative 7; Mismatches 3; Indels 8; Gaps 2;
QY 1 MELAAICRWGILLALLPFGAASVCTGTDMKILRLPASPEFHLMRLHYGCGVQGNL 60
DB 1 MELAAICRWGILLALLPFGAASVCTGTDMKILRLPASPEFHLMRLHYGCGVQGNL 60
QY 61 ELTYLPTNASLSFLDIOEVQGVYLIANQVRQVPLQRLIRVGTQLFEDNVYALAVLDNG 120
DB 61 ELTYLPTNASLSFLDIOEVQGVYLIANQVRQVPLQRLIRVGTQLFEDNVYALAVLDNG 120
QY 121 DPLNNTPTVTGASPGGLBELQRLSTELIKGGLVLIQRNPOLCYODTILMKDIFKHNOILA 180
DB 121 DPLNNTPTVTGASPGGLBELQRLSTELIKGGLVLIQRNPOLCYODTILMKDIFKHNOILA 180
QY 181 LTLIDTNRBRACHPQSPCKGSRGSESDCOSLTRVYAGGACRCKGRLPTCCHEQC 240
DB 181 LTLIDTNRBRACHPQSPCKGSRGSESDCOSLTRVYAGGACRCKGRLPTCCHEQC 240
QY 241 AAGCTGPKASDCLAHFNHSGICELHCPALVYNTDTFESMPNBEGRYTFGASCVTACP 300
DB 241 AAGCTGPKASDCLAHFNHSGICELHCPALVYNTDTFESMPNBEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPRLHNOEVTADGTQCEKCKSPCARVCYGLGMEH-----LREV 353
DB 301 YNYLSTDVGSCTLVCPRLHNOEVTADGTQCEKCKSPCARVCYGLGMEH-----LREV 353
QY 360 PKVASHLBEFAGCKKIFGSLAFLEPESFGDPASNTAPLQPEQLQVFETLEITGYLYIS 419
DB 360 PKVASHLBEFAGCKKIFGSLAFLEPESFGDPASNTAPLQPEQLQVFETLEITGYLYIS 419
QY 420 AMPDSLPLDSVFQNTQVIRGRILNNGAVSLTLQGLGISWLGRLSRLGSLALIHNTHT 479
DB 420 AMPDSLPLDSVFQNTQVIRGRILNNGAVSLTLQGLGISWLGRLSRLGSLALIHNTHT 479
QY 480 LCFVHTVPMDOQLFRPHQALLHTARPEDECVGEGELACHQICARGHCMGPPTQCVNCSQ 539
DB 480 LCFVHTVPMDOQLFRPHQALLHTARPEDECVGEGELACHQICARGHCMGPPTQCVNCSQ 539
QY 474 LCFVHTVPMDOQLFRPHQALLHTARPEDECVGEGELACHQICARGHCMGPPTQCVNCSQ 533
DB 474 LCFVHTVPMDOQLFRPHQALLHTARPEDECVGEGELACHQICARGHCMGPPTQCVNCSQ 533
QY 540 FLRGQECVBEKCVLQGLPREVYNABHCLPCHBECQPOGNSVTCFGEPAADQCVAAHYKDP 599
DB 540 FLRGQECVBEKCVLQGLPREVYNABHCLPCHBECQPOGNSVTCFGEPAADQCVAAHYKDP 599
QY 534 FLRGQECVBEKCVLQGLPREVYNABHCLPCHBECQPOGNSVTCFGEPAADQCVAAHYKDP 593
DB 534 FLRGQECVBEKCVLQGLPREVYNABHCLPCHBECQPOGNSVTCFGEPAADQCVAAHYKDP 593
QY 600 PFCVAPCPGSGVPRDLSYPIWKFPEBEBAQCPCLNCHNSCVLDLDDKCPAEQASPLTS 659
DB 600 PFCVAPCPGSGVPRDLSYPIWKFPEBEBAQCPCLNCHNSCVLDLDDKCPAEQASPLTS 659
QY 660 IVSAVVGILLVVVLGVFGILLIKRQOKIRKXTYMRLLQETELVEPLTPSGAMPNOQMR 719
DB 654 IVSAVVGILLVVVLGVFGILLIKRQOKIRKXTYMRLLQETELVEPLTPSGAMPNOQMR 713
QY 720 ILKETELRKVKVLSGAFGTGKGIWIPGENVKIIPVAIKVLRNTSPKANKELIDEAYV 779

```


Db 714 ILKTELKRVKVLGSGAGFTYKGIWIDGENVKI PVAIKYLRNTSPKANKELIDEAYV 773
 Qy 780 MAGVSPYVSRLLGICLTSTVOLVTLQMLPYGCLDHYENRGRLGSDLLMWCQIAKGM 839
 Db 774 MAGVSPYVSRLLGICLTSTVOLVTLQMLPYGCLDHYENRGRLGSDLLMWCQIAKGM 833
 Qy 840 STLEDVRLVHRDLAARNVLVSPNWKITDGLARLLDIDETEVHADGKVPIMMALES 899
 Db 834 STLEDVRLVHRDLAARNVLVSPNWKITDGLARLLDIDETEVHADGKVPIMMALES 893
 Qy 900 ILRRFTHQSDVWSGVYTWELMTFGAKPYDGIPIAREIPDLLEXGERLPOPICTIDVYM 959
 Db 894 ILRRFTHQSDVWSGVYTWELMTFGAKPYDGIPIAREIPDLLEXGERLPOPICTIDVYM 953
 Qy 960 IWKCMIDSECRPRPRELVSFSFMSAADPQRFVVIQNEGLGAPSPLDSTFYRSLEDD 1019
 Db 954 IWKCMIDSECRPRPRELVSFSFMSAADPQRFVVIQNEGLGAPSPLDSTFYRSLEDD 1013
 Qy 1020 MGDLDAAEELVLPQGGFCPPDAPAGGMVHRRSSSTRSGGDLTLGLEPSEEARPS 1079
 Db 1014 MGDLDAAEELVLPQGGFCPPDAPAGGMVHRRSSSTRSGGDLTLGLEPSEEARPS 1073
 Qy 1080 PLAPSEGAGSDVFDGDLGMAKAGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYVAPLTC 1139
 Db 1074 PLAPSEGAGSDVFDGDLGMAKAGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYVAPLTC 1133
 Qy 1140 SPQPEYVNOPDVRPQPPSPREGLPLAARPAATLBRATKLSPGKNGVYKDVFAFGAVEN 1199
 Db 1134 SPQPEYVNOPDVRPQPPSPREGLPLAARPAATLBRATKLSPGKNGVYKDVFAFGAVEN 1193
 Qy 1200 PEYLTPOGGAAPQPPPAFSPAFNLLYYWODDPERGAPSTFKGPTAENPEYGLDY 1259
 Db 1194 PEYLTPOGGAAPQPPPAFSPAFNLLYYWODDPERGAPSTFKGPTAENPEYGLDY 1253
 Qy 1260 PY 1261
 Db 1254 PY 1255
 Db
 RESULT 4
 AAU74545
 ID AAU74545 standard; Protein; 1255 AA.
 AC AAU74545;
 XX
 DT 23-APR-2002 (first entry)
 XX
 DE Human HER2 (ErbB2) polypeptide.
 XX
 KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
 KW glial disorder; astrocytal disorder; hypothalamic disorder;
 KW glandular disorder; macropneagal disorder; epithelial disorder;
 KW stromal disorder; blastococelic disorder; inflammatory disorder;
 KW angiogenic disorder; immunological disorder.
 XX
 OS Homo sapiens.
 XX
 PN US:002001587-A1.
 XX
 PD 03-JAN-2002.
 XX
 PF 16-MAR-2001; 2001US-0811123.
 XX
 PR 16-MAR-2000; 2000US-189844P.
 PR 05-OCT-2000; 2000US-238927P.
 XX
 PA (ERIC/) ERICKSON S.
 PA (SCHW/) SCHWALL R.
 PA (SLIW/) SLIWOMSKI M.
 XX

PI Erickson S, Schwall R, Sliwowski M;
 XX
 DR MPI: 2002-163686/21.
 DR N-PSDB; ABR14058.
 XX
 PT Treating tumour characterised by overexpression of epidermal growth
 factor receptor, ErbB or cancer in mammal, comprises administering
 anti-ErbB antibody-maytansinoid conjugate to the mammal
 XX
 PS Example 3; Fig 7; 93pp; English.
 CC The invention relates to treating a tumour in a mammal, where the tumour
 CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (ErbB) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses ErbB2. The method is also useful for treating
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macropneagal,
 CC epithelial, stromal, blastococelic, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)
 CC polypeptide of the invention.
 XX
 SQ Sequence 1255 AA;
 Query Match 98.3%; Score 6736; DB 23; Length 1255;
 Best Local Similarity 98.6%; Pred. No. 0;
 Matches 1244; Conservative 7; Mismatches 3; Indels 8; Gaps 1;
 Qy 1 MELALCRWGLLALLPFGAASVCTGTDMKRLPASPETHLDMRLHYQCGVQGNL 60
 Db 1 MELALCRWGLLALLPFGAASVCTGTDMKRLPASPETHLDMRLHYQCGVQGNL 60
 Qy 61 ELTYLPTNASSFLDIOEVGVYLLAHNGVQVPLQRLAIYRGTQLFEDNYALAVLDNG 126
 Db 61 ELTYLPTNASSFLDIOEVGVYLLAHNGVQVPLQRLAIYRGTQLFEDNYALAVLDNG 126
 Qy 121 DPLNNTPTVGTASFGSLRELOLRSTELIKGVLIGRNPOLCYQDITLMMDIFHKNNQLA 186
 Db 121 DPLNNTPTVGTASFGSLRELOLRSTELIKGVLIGRNPOLCYQDITLMMDIFHKNNQLA 186
 Qy 181 LTLIDTNSRACHPCSPCKGSRGWSSESDCSLITFVCAAGGACARCKGPLPTDCHEQC 246
 Db 181 LTLIDTNSRACHPCSPCKGSRGWSSESDCSLITFVCAAGGACARCKGPLPTDCHEQC 246
 Qy 241 AAGCTGPRGSDCLCLPHNNSGICELCPALVTYNTDTEFSMNPBGRYFFGASCVTACP 306
 Db 241 AAGCTGPRGSDCLCLPHNNSGICELCPALVTYNTDTEFSMNPBGRYFFGASCVTACP 306
 Qy 301 YNYLSTVGSCTVLCPLHNEVTAEDGTORCEKSKPCARVCYGLGFMNFVFSFWLR-V 359
 Db 301 YNYLSTVGSCTVLCPLHNEVTAEDGTORCEKSKPCARVCYGLGFMNFVFSFWLR-V 359
 Qy 354 PKVASHLEFRAGCKKIFGSLAFIPSEFDGDPASNTAPLPDQQLVFEETLEETIGLYLIS 419
 Db 354 RAVTSANIQEFAGCKKIFGSLAFIPSEFDGDPASNTAPLPDQQLVFEETLEETIGLYLIS 413
 Qy 420 AMPDLPDLISVFQMLQVIRGRILHNGAYSLTLQGLISWLGRLSELGSLLIHHNNH 479
 Db 420 AMPDLPDLISVFQMLQVIRGRILHNGAYSLTLQGLISWLGRLSELGSLLIHHNNH 473
 Qy 474 LCFVHTVPMQDLFFNPQALHHTANREDECEVEGGLACHQLCARGHCWGGPTQVNCSCQ 533
 Db 474 LCFVHTVPMQDLFFNPQALHHTANREDECEVEGGLACHQLCARGHCWGGPTQVNCSCQ 533
 Qy 540 FLRGQCEVEECRYVQGLPREYVNAARHCLPCHPECOQNGSVTFCGFEADQCAVCAHYKXP 599
 Db 534 FLRGQCEVEECRYVQGLPREYVNAARHCLPCHPECOQNGSVTFCGFEADQCAVCAHYKXP 593
 Qy 600 PFVACRCPGVKPDLSYMPIMKPEDEGACQPCPINCTHSCVDLDDKGGPAPGASPLTS 659

```

Db      594 PFCVARSQGVKPDLSYBIMKFPDEEACOPCPINCHSCVDLDDGCPDEQASLTS 653
Qy      660 TYSAVGILLVVLGVVGIILIKRQOKIRKYTRMRILQETELVEPLTPSGAMPNQAMR 719
Db      654 TYSAVGILLVVLGVVGIILIKRQOKIRKYTRMRILQETELVEPLTPSGAMPNQAMR 713
Qy      720 ILKETELRKVKVLSGSAFGTVYKGIWIPDGENVKI PVAIKVLRNTSPKANKEILDEAYV 779
Db      714 ILKETELRKVKVLSGSAFGTVYKGIWIPDGENVKI PVAIKVLRNTSPKANKEILDEAYV 773
Qy      780 MAGVSPYVSRLLGLITLTSTVQLVQLMVEYGLDHNVENGRIGSDQLMWCQIAXKM 839
Db      774 MAGVSPYVSRLLGLITLTSTVQLVQLMVEYGLDHNVENGRIGSDQLMWCQIAXKM 833
Qy      840 SYLEDVRLVARDIAANVLVKSPPNHVKTDFGLARLIDIDETEVADGKVPIMKMALES 899
Db      834 SYLEDVRLVARDIAANVLVKSPPNHVKTDFGLARLIDIDETEVADGKVPIMKMALES 893
Qy      900 ILRRRFTHOSDVSYGVTWELMTFGAKYDGIIPAREIPDLLEKGERLPOPICITIDVYM 959
Db      894 ILRRRFTHOSDVSYGVTWELMTFGAKYDGIIPAREIPDLLEKGERLPOPICITIDVYM 953
Qy      960 IMVCKMWIDECRPRRELVSERSRMAPDQRFVYIQNEDLGASPLDSTFYRSLLEDD 1019
Db      954 IMVCKMWIDECRPRRELVSERSRMAPDQRFVYIQNEDLGASPLDSTFYRSLLEDD 1013
Qy      1020 MGDLYDAEYLVPOQGFPCDPAPGAGVNRHRSSTRSGGGLTIGLEPSEEAR 1079
Db      1014 MGDLYDAEYLVPOQGFPCDPAPGAGVNRHRSSTRSGGGLTIGLEPSEEAR 1073
Qy      1080 PLASEGASDVDPDGLMGAAKGLSLPTHPSPLORESSEPTVPLPSETDGYAPLTC 1139
Db      1074 PLASEGASDVDPDGLMGAAKGLSLPTHPSPLORESSEPTVPLPSETDGYAPLTC 1133
Qy      1140 SPOPEYVNPDPVPPPPSPREGPLPAPAPAGATLEBAKTLSPQKGVVQVDAFAGAVEN 1199
Db      1134 SPOPEYVNPDPVPPPPSPREGPLPAPAPAGATLEBAKTLSPQKGVVQVDAFAGAVEN 1193
Qy      1200 PEYLTPGGGAAPQHPAPSPAFDLVYWDQPPRERGAPSTFGKTPFAENPEYLGLDV 1259
Db      1194 PEYLTPGGGAAPQHPAPSPAFDLVYWDQPPRERGAPSTFGKTPFAENPEYLGLDV 1253
Qy      1260 PV 1261
Db      1254 PV 1255

RESULT 5
AAW01111 standard; Protein; 1255 AA.
ID      AAW01111
AC      AAW01111;
DT      01-JAN-1997 (first entry)
XX
DE      HER-2/neu protein.
XX
KM      HER-2/neu: c-erbB1; p185; oncogene; tyrosine protein kinase;
KM      breast cancer; ovary cancer; colon cancer; lung cancer;
KM      prostate cancer; immunisation; tumour; vaccine; vector.
OS      Homo sapiens.
XX
XX
XX      Key      Location/Qualifiers
XX      FT      676..1255
XX      Domain /label= Intracellular domain
XX      FT      /note= "claimed domain, useful for immunisation"
XX
XX      W09430514-A1.
XX      03-OCT-1996.
XX

```

```

PF      28-MAR-1996; 96WO-US01689.
XX
XX      31-MAR-1995; 95US-0414417.
PR
XX      (UNIM ) UNIV WASHINGTON.
XX
XX      Cheever MA, Disis ML;
PI      WPI, 1996-455361/45.
XX      N-PSDB; AAT40739.
XX
XX      DNA encoding HER-2-neu polypeptide(s) - used for prevention or
PT      treatment of malignancies with which the HER-2/neu oncogene is
PT      associated
XX
XX      Claim 2; Page 56-61; 71pp; English.
XX
XX      Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC      the product of the HER-2/neu oncogene (see also AAT40739). The
CC      protein is over-expressed in various cancers, including breast,
CC      ovarian, colon, lung and prostate. The intracellular domain of the
CC      protein can be used to immunise an animal against a malignancy with
CC      which the oncogene is associated. The polypeptide can be produced
CC      in transformed host cells for use in immunisation. Alternatively,
CC      animal cells are transfected in vivo or ex vivo with a viral vector
CC      that directs expression of the polypeptide.
XX
SQ      Sequence 1255 AA:
XX
XX      Query Match 98.2%; Score 6730; DB 17; Length 1255;
XX      Best Local Similarity 98.4%; Pred. No. 0;
XX      Matches 1242; Conservative 8; Mismatches 4; Indels 8; Gaps 2.
Qy      1 MELAAACRAGLLALLPAGASTQVCTGDMKLRLPASPETHLDMLRHLVYGCQVVGNTL 60
Db      1 MELAAACRAGLLALLPAGASTQVCTGDMKLRLPASPETHLDMLRHLVYGCQVVGNTL 60
Qy      61 ELTYLPNLSLFLQDIOGVGVYLIANQVRQVPLORLIRYRGQLVEDNYALAVLDNG 120
Db      61 ELTYLPNLSLFLQDIOGVGVYLIANQVRQVPLORLIRYRGQLVEDNYALAVLDNG 120
Qy      121 DPLNNTPTVPGASPGGLRELQRLSLTEILKGVLIQRNPQLCYQDTILMKDIFHNQOLA 180
Db      121 DPLNNTPTVPGASPGGLRELQRLSLTEILKGVLIQRNPQLCYQDTILMKDIFHNQOLA 180
Qy      181 LTLIDNRSRACHPCSPMCKGRSCWGESSEDCSLTRVYACGACRCAGPLPTDCHEQC 240
Db      181 LTLIDNRSRACHPCSPMCKGRSCWGESSEDCSLTRVYACGACRCAGPLPTDCHEQC 240
Qy      241 AAGCTGPKASDCLACLHPNHSICELHCAALTYNMTFESMPNDEGRYTFGASCVTACP 300
Db      241 AAGCTGPKASDCLACLHPNHSICELHCAALTYNMTFESMPNDEGRYTFGASCVTACP 300
Qy      301 YNYLSTVSGCTLVCPAHNOEYVAEDGTORCEKSKPCARVCGYGLGMFNNFTVSFWLR-V 359
Db      301 YNYLSTVSGCTLVCPAHNOEYVAEDGTORCEKSKPCARVCGYGLGMFNNFTVSFWLR-V 359
Qy      360 PKVASHLPEFPACCKIFGSLALPSSFGDPASNTAPIQEPOLQVFEETLEITGLVYS 419
Db      354 RAYTSANIOFAGCKKIFGSLALPSSFGDPASNTAPIQEPOLQVFEETLEITGLVYS 413
Qy      420 AMPSLPDLSPVQNLVYRIRIIRHNGVSYLTLOGLSIMLGLRSLRELSGLALIHNNH 479
Db      414 AMPSLPDLSPVQNLVYRIRIIRHNGVSYLTLOGLSIMLGLRSLRELSGLALIHNNH 473
Qy      480 LCFVHTVPMQDLFRNPQALLTANRDEDECVESGLAQQLCARHGCKWPGPTQVCNQG 539
Db      474 LCFVHTVPMQDLFRNPQALLTANRDEDECVESGLAQQLCARHGCKWPGPTQVCNQG 533
Qy      540 FLRGQCEVEECRYLQGLPREYVNAARHCLPCHPCQOPNGSVTCFGEAQQCYACATYKXP 599
Db      534 FLRGQCEVEECRYLQGLPREYVNAARHCLPCHPCQOPNGSVTCFGEAQQCYACATYKXP 593

```

QY	600	PFCCARPGSGVKDLSYMPIMKFPDEEGACQCPINCHSGVLDLDDGCAEORAPLTS	650
Db	594	PFCARPGSPVKRDLSTYMPIKFPDEEGACQCPINCHSGVLDLDDGCAEORAPLTS	653
QY	660	IVSAVGIILLVVLGVVFGIILKRQOKIRKYTNRRLLQETELVEPTPSGAMPNOQR	719
Db	654	IISAVGILLVVLGVVFGIILKRQOKIRKYTNRRLLQETELVEPTPSGAMPNOQR	713
QY	720	ILKETELIRKYKLGSGAFQTVYKGIWIPDGEVVKIPVAIKYLRNTPSPKANKETIDEAVY	779
Db	714	ILKETELIRKYKLGSGAFQTVYKGIWIPDGEVVKIPVAIKYLRNTPSPKANKETIDEAVY	773
QY	780	MAGVSPYVSRLIGICLTSTVQLVITOLMPYCGLLDHYENNGRIGSODLLNMCQIAKGM	839
Db	774	MAGVSPYVSRLIGICLTSTVQLVITOLMPYCGLLDHYENNGRIGSODLLNMCQIAKGM	833
QY	840	SYLEDVAVLHRDLAARNVLYKSPNHVKITDGLARLLDIDETVHADQGVPIKMALES	899
Db	834	SYLEDVAVLHRDLAARNVLYKSPNHVKITDGLARLLDIDETVHADQGVPIKMALES	893
QY	900	ILRRRFTHQSDVMSYGVITWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM	959
Db	894	ILRRRFTHQSDVMSYGVITWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM	953
QY	960	IMVYCMWIDSECRPRERELVSSEFSKVARDPQRFVVIQNEDLGPASPLDSTFYRSLLDEDD	1013
Db	954	IMVYCMWIDSECRPRERELVSSEFSKVARDPQRFVVIQNEDLGPASPLDSTFYRSLLDEDD	1011
QY	1020	MGLDVAAEEVLVQOGSFCCPPDPAFAGAGMHHRRSSSTRGGGDLTGLBSEEEAPRS	1077
Db	1014	MGLDVAAEEVLVQOGSFCCPPDPAFAGAGMHHRRSSSTRGGGDLTGLBSEEEAPRS	1073
QY	1080	PLAPSEGAGSDVVDGDLGMAAKGLQSLPTHDPSPLOKRSYSDPVPVPSRTDGYVAPLTC	1133
Db	1074	PLAPSEGAGSDVVDGDLGMAAKGLQSLPTHDPSPLOKRSYSDPVPVPSRTDGYVAPLTC	1133
QY	1140	SPOPEYVNOGDVAFPOPPSPREBQPLPAAPPAQATLIERAKTSSPGKNGVVKQVFAFGAVEN	1199
Db	1134	SPOPEYVNOGDVAFPOPPSPREBQPLPAAPPAQATLIERAKTSSPGKNGVVKQVFAFGAVEN	1193
QY	1200	PEVLTPOGGAAPQHPHPAPFSPAFPNLLYYMDDPPREGAPSTFKGPTANPEVYLGLDV	1255
Db	1194	PEVLTPOGGAAPQHPHPAPFSPAFPNLLYYMDDPPREGAPSTFKGPTANPEVYLGLDV	1255
QY	1260	PV 1261	
Db	1254	PV 1255	
RESULT 6			
AAW92406			
ID	AAW92406 standard; Protein; 1255 AA.		
XX	AAW92406;		
XX	21-APR-1999 (first entry)		
XX	Human HER-2/neu oncogene protein.		
XX	HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;		
XX	malignancy; treatment; tumour.		
XX	Homo sapiens.		
XX	Key		
XX	Region		
XX	Location/Qualifiers		
XX	676..1255		
XX	/note="region which elicits immune response"		
XX	US5869445-A.		
XX	09-FEB-1999.		
XX	01-APR-1996; 96US-0625101.		

XX	01-APR-1996;	9605-0625101.
PR	17-MAR-1993;	9305-0038644.
PR	12-AUG-1993;	9305-0106112.
PR	31-MAR-1995;	9505-0414417.
XX		
PA	(UNIM) UNIV WASHINGTON.	
XX		
PI	Cheever MA, Disis ML;	
XX		
DR	WPI, 1999-152835/13.	
XX	N-PSDB; AAX01912.	
PT	Use of HER-2/neu polypeptides - for eliciting an immune response to	
XX	an HER-2/neu associated malignancy, particularly for treating or	
XX	preventing tumours	
XX		
PS	Claim 3, Column 31-38; 26pp; English.	
XX		
CC	This sequence represents the human HER-2/neu oncogene protein. A fragmen	
CC	of this protein is used in a method for eliciting or enhancing an immune	
CC	response to HER-2/neu protein. The polypeptide can stimulate T cells and	
CC	B cells to produce an immune response to the HER-2/neu protein. The	
CC	method can be used for immunisation against a malignancy in which the	
CC	HER-2/neu oncogene is associated and in the treatment of an existing	
CC	tumour, or to prevent tumour occurrence or reoccurrence.	
XX		
XX	Sequence 1255 AA:	
Query Match	98.2%; Score 6730; DB 20; Length 1255;	
Best Local Similarity	98.4%; Pred. No. 0;	
Matches 1244;	Conservative 8; Mismatches 4; Indels 8; Gaps	
QY	1 METAAALCRWGLLALLPPGAASTQVCTGDMKRLRPASPEHLMDRLYQGCQVQGNL	60
Db	1 METAAALCRWGLLALLPPGAASTQVCTGDMKRLRPASPEHLMDRLYQGCQVQGNL	60
QY	61 ELYLYLPTNASLSLQDICEQGVVLAHNOVQVLORLIVRGQLFEDNYALAVLDNG	120
Db	61 ELYLYLPTNASLSLQDICEQGVVLAHNOVQVLORLIVRGQLFEDNYALAVLDNG	120
QY	121 DPINNTPTVYASPGGLRELOLRSLTEILKGGVLLRNRPOLCYQDTIMKDIPIKNNOLA	180
Db	121 DPINNTPTVYASPGGLRELOLRSLTEILKGGVLLRNRPOLCYQDTIMKDIPIKNNOLA	180
QY	181 LTLIDNBRARACPGSPCKMGSGRSGESSEDDQSLTRIVCAGGCAKCGPLPTCCHEQC	240
Db	181 LTLIDNBRARACPGSPCKMGSGRSGESSEDDQSLTRIVCAGGCAKCGPLPTCCHEQC	240
QY	241 AAGCTPCKNSDCLACLFHNSGICGLHPALVTYNTDFFSEMPNDEGRVYTGASCVTACP	300
Db	241 AAGCTPCKNSDCLACLFHNSGICGLHPALVTYNTDFFSEMPNDEGRVYTGASCVTACP	300
QY	301 YNLTSDVNSCTLVGRILHNOETIADTQRCCKSKPCARVCYGLGEMNNNTVSFWLR-V	360
Db	301 YNLTSDVNSCTLVGRILHNOETIADTQRCCKSKPCARVCYGLGEMNNNTVSFWLR-V	360
QY	360 PKYSASHLEEFAGCKKIFGSLAFLESEFDPASNTAPALQPECLQVETLEITGYLYIS	420
Db	360 PKYSASHLEEFAGCKKIFGSLAFLESEFDPASNTAPALQPECLQVETLEITGYLYIS	420
QY	420 AWPDSLPDLISVQNLQVIRGRIILHNGAVSLTLOGLSWGLRSRLRELGSGLAIINHTH	480
Db	420 AWPDSLPDLISVQNLQVIRGRIILHNGAVSLTLOGLSWGLRSRLRELGSGLAIINHTH	480
QY	480 LCFVHTVPMDOQLFRPHQALLHTARPEDECVGEGGLACHQICARGHCMGPPPTCVVCSQ	540
Db	480 LCFVHTVPMDOQLFRPHQALLHTARPEDECVGEGGLACHQICARGHCMGPPPTCVVCSQ	540
QY	540 FLRGQCEVCECVLQGLPREYVNAHCLPCRPCEQOPONGSYTCGPRADQVCAAHYKDP	600
Db	540 FLRGQCEVCECVLQGLPREYVNAHCLPCRPCEQOPONGSYTCGPRADQVCAAHYKDP	600
QY	600 FLRGQCEVCECVLQGLPREYVNAHCLPCRPCEQOPONGSYTCGPRADQVCAAHYKDP	660
Db	600 FLRGQCEVCECVLQGLPREYVNAHCLPCRPCEQOPONGSYTCGPRADQVCAAHYKDP	660

```

QY 600 PFCVACPSGVKPDLSYMPIMKFPDEGACQPCPINCTHSCVDLDKCKGAPAEQASPLTS 659
DB 594 PFCVACPSGVKPDLSYMPIMKFPDEGACQPCPINCTHSCVDLDKCKGAPAEQASPLTS 653
QY 660 TSVAVVGIILVTVLVGVVFGILLKRRQOKTRKTYMRLLQETELVEPLTPSGAMPNQAQR 719
DB 654 IISAVVGIILVTVLVGVVFGILLKRRQOKTRKTYMRLLQETELVEPLTPSGAMPNQAQR 713
QY 720 ILKETELRKVKVLGSAFGVYVKGIIIPGENVKIPVAIKVIRENTSKRANKELIDEXYV 779
DB 714 ILKETELRKVKVLGSAFGVYVKGIIIPGENVKIPVAIKVIRENTSKRANKELIDEXYV 773
QY 780 MAGVSPVYSRLIGILITSTVQLVTLQMLPBYGCLLDHVENRGRIGSODLLMCMQIAKM 839
DB 774 MAGVSPVYSRLIGILITSTVQLVTLQMLPBYGCLLDHVENRGRIGSODLLMCMQIAKM 833
QY 840 SYLEDVRLVHRDLAARNVLKSPNHVKINDPGLALDIDETEVHADGKXPIKMALES 899
DB 834 SYLEDVRLVHRDLAARNVLKSPNHVKINDPGLALDIDETEVHADGKXPIKMALES 893
QY 900 ILRRFTHQS DVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYM 959
DB 894 ILRRFTHQS DVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYM 953
QY 960 IMVKCWMIDSECHPRRELVSFSRMAARDPQRFVVIQNEIDGFPASPLDSTFYRSLIEDDD 1019
DB 954 IMVKCWMIDSECHPRRELVSFSRMAARDPQRFVVIQNEIDGFPASPLDSTFYRSLIEDDD 1013
QY 1020 MGDVDAEEYLVPOQGFPCPDPAFGAGVHHRRSSSTRSGGDLTLGLPSEEBAPRS 1079
DB 1014 MGDVDAEEYLVPOQGFPCPDPAFGAGVHHRRSSSTRSGGDLTLGLPSEEBAPRS 1073
QY 1080 PLAPSEGASDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTC 1139
DB 1074 PLAPSEGASDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTC 1133
QY 1140 SPOPEYVNOPDVRPOPSPREGPLPAPAPAGATLERAKTLSGKXGVKDYAFAGCAVEN 1199
DB 1134 SPOPEYVNOPDVRPOPSPREGPLPAPAPAGATLERAKTLSGKXGVKDYAFAGCAVEN 1193
QY 1200 PEYLTFOGGAAPQHPAPAFSPAFLNLYWMDPPBERGAPSTFGKPTAENPEYLGIDV 1259
DB 1194 PEYLTFOGGAAPQHPAPAFSPAFLNLYWMDPPBERGAPSTFGKPTAENPEYLGIDV 1253
QY 1260 PV 1261
DB 1254 PV 1255

```

RESULT 7
 ID AAB21198 standard: protein; 1255 AA.
 XX AAB21198;

DT 12-JAN-2001 (first entry)
 XX Human HER-2/neu protein.
 DE Human HER-2/neu protein.
 XX Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
 KM breast cancer; prostate cancer; ovarian cancer; lung cancer;
 KW colon cancer;
 XX Homo sapiens.
 OS Homo sapiens.
 PN MO2000044899-A1.
 XX MO2000044899-A1.
 PD 03-AUG-2000.
 XX 03-AUG-2000.
 PF 28-JAN-2000; 2000MO-US02164.
 XX 28-JAN-2000; 2000MO-US02164.
 PR 29-JAN-1999; 99US-0117976.
 XX 29-JAN-1999; 99US-0117976.

```

PA (CORI-) CORIXA CORP.
PA (SMIK) SMITHKLINE BEECHAM.
PI Cheever MA, Ghyesen D;
XX WPI; 2000-505976/45.
DR N-PSDB; AAA89736.
XX
PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers.
XX
PS Claim 52; Fig 7; 128pp; English.
XX
CC The present sequence is the human HER-2/neu protein. It is a member of
CC the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against
CC these neoplasias.
XX
SQ Sequence 1255 AA;

```

Query Match 98.2%; Score 6730; DB 21; Length 1255;
 Best Local Similarity 98.4%; Pred. No. 0;
 Matches 1242; Conservative 8; Mismatches 4; Indels 8; Gaps

```

QY 1 MELALACRMGILLALPPGAASVQVCTGDMKRLPASPETHLMDRLHYGGCQWQGNL 60
DB 1 MELALACRMGILLALPPGAASVQVCTGDMKRLPASPETHLMDRLHYGGCQWQGNL 60
QY 61 ELITYPTNASSLFDLIDVQGVYIIAHNVQVPLQRLIVRGTQLFEDNYALAVLDNG 120
DB 61 ELITYPTNASSLFDLIDVQGVYIIAHNVQVPLQRLIVRGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTPTVAGSPGGLRELOSLTELKXGVLQIONPOLCYODTLIMQDIFKXNOLA 180
DB 121 DPLNNTPTVAGSPGGLRELOSLTELKXGVLQIONPOLCYODTLIMQDIFKXNOLA 180
QY 181 LTLIDNRSRACHPCSPKCKSGRSGESSEDCQSLTRVACAGCARGKGLPTDCHEQC 240
DB 181 LTLIDNRSRACHPCSPKCKSGRSGESSEDCQSLTRVACAGCARGKGLPTDCHEQC 240
QY 241 AAGCTGPRHSDCLALHNHSGICELHPALVYNTDTFESMNPDEGRYTFGASCVTACP 300
DB 241 AAGCTGPRHSDCLALHNHSGICELHPALVYNTDTFESMNPDEGRYTFGASCVTACP 300
QY 301 YNYLSTDVSGCTVCLPLHNOEVTAEADGTORCEKSKPCARVCYGLGMFNNPTVFWLR-V 350
DB 301 YNYLSTDVSGCTVCLPLHNOEVTAEADGTORCEKSKPCARVCYGLGMFNNPTVFWLR-V 350
QY 350 YNYLSTDVSGCTVCLPLHNOEVTAEADGTORCEKSKPCARVCYGLGMFNNPTVFWLR-V 350
DB 350 YNYLSTDVSGCTVCLPLHNOEVTAEADGTORCEKSKPCARVCYGLGMFNNPTVFWLR-V 350
QY 360 PKVSAHLEEFPAQCKKIFGSLAPLPESFDGDPASNTAPLPQEQLOVEETLEITGYLYIS 410
DB 360 PKVSAHLEEFPAQCKKIFGSLAPLPESFDGDPASNTAPLPQEQLOVEETLEITGYLYIS 410
QY 354 RAVTSANIOEFAGCKKIFGSLAPLPESFDGDPASNTAPLPQEQLOVEETLEITGYLYIS 413
DB 354 RAVTSANIOEFAGCKKIFGSLAPLPESFDGDPASNTAPLPQEQLOVEETLEITGYLYIS 413
QY 420 AMPDLPPLSVFQNLQVGRILHNHGSLSLTLQGLGISMGLSLRLREGLALLIHNNTH 475
DB 420 AMPDLPPLSVFQNLQVGRILHNHGSLSLTLQGLGISMGLSLRLREGLALLIHNNTH 475
QY 475 AMPDLPPLSVFQNLQVGRILHNHGSLSLTLQGLGISMGLSLRLREGLALLIHNNTH 475
DB 475 AMPDLPPLSVFQNLQVGRILHNHGSLSLTLQGLGISMGLSLRLREGLALLIHNNTH 475
QY 480 LCFVHTVPMQDLFRNPHQALHTANREDECVGEGALCHQLCARGHMGWGPPTQVNNCSQ 539
DB 480 LCFVHTVPMQDLFRNPHQALHTANREDECVGEGALCHQLCARGHMGWGPPTQVNNCSQ 539
QY 539 LCFVHTVPMQDLFRNPHQALHTANREDECVGEGALCHQLCARGHMGWGPPTQVNNCSQ 539
DB 539 LCFVHTVPMQDLFRNPHQALHTANREDECVGEGALCHQLCARGHMGWGPPTQVNNCSQ 539
QY 540 FLRGOECVEECRVLOGLPREYVNAHCLPCHPPECQPNQSVTTCGPPADQCVACAHYKCP 599
DB 540 FLRGOECVEECRVLOGLPREYVNAHCLPCHPPECQPNQSVTTCGPPADQCVACAHYKCP 599
QY 599 FLRGOECVEECRVLOGLPREYVNAHCLPCHPPECQPNQSVTTCGPPADQCVACAHYKCP 599
DB 599 FLRGOECVEECRVLOGLPREYVNAHCLPCHPPECQPNQSVTTCGPPADQCVACAHYKCP 599
QY 600 PFCVACPSGVKPDLSYMPIMKFPDEGACQPCPINCTHSCVDLDKCKGAPAEQASPLTS 659
DB 600 PFCVACPSGVKPDLSYMPIMKFPDEGACQPCPINCTHSCVDLDKCKGAPAEQASPLTS 659

```

```

Db      534 PFCVACRCSGVKPDLSIYPIKPFDEBEGACOPCLNCTHSCVLDLDDGCAEERASPLTS 653
Qy      660 IVSAVVGILLVVLGVVEGILLIKRQCKIRKRYTRRLLOETELVEPLTPSGAMPNQAKR 719
Db      654 IISAVVGILLVVLGVVEGILLIKRQCKIRKRYTRRLLOETELVEPLTPSGAMPNQAKR 713
Qy      720 ILKETELRKVKVLGSGAGFTYKGIWTDGNNVKI PVAIKYLRNTSPKANKETLDEAYV 779
Db      714 ILKETELRKVKVLGSGAGFTYKGIWTDGNNVKI PVAIKYLRNTSPKANKETLDEAYV 773
Qy      780 MAGVSPYVSRLLGICTSTVOLVTQLMPCYCLLDHYRENGRGRSGODLNMCKQIAKGM 839
Db      774 MAGVSPYVSRLLGICTSTVOLVTQLMPCYCLLDHYRENGRGRSGODLNMCKQIAKGM 833
Qy      840 SYLEDVRLVHRDLAARNVLVSPNHVKITDFGLARLLDIDETEHADGKVPYKMMALES 899
Db      834 SYLEDVRLVHRDLAARNVLVSPNHVKITDFGLARLLDIDETEHADGKVPYKMMALES 893
Qy      900 ILRRFTHQSDVWSYGVTVWELMTFGAKPYGIPAREIPOLLEKGEPLPOPICTIDVTM 959
Db      894 ILRRFTHQSDVWSYGVTVWELMTFGAKPYGIPAREIPOLLEKGEPLPOPICTIDVTM 953
Qy      960 IMVKCMIDSECRPFRELVSSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDD 1019
Db      954 IMVKCMIDSECRPFRELVSSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDD 1013
Qy      1020 MGDLYDAEYLYVPOGFCPPDPAAGAGMHHRRSSSTSGGDLTLGLEPSEEBAPRS 1079
Db      1014 MGDLYDAEYLYVPOGFCPPDPAAGAGMHHRRSSSTSGGDLTLGLEPSEEBAPRS 1073
Qy      1080 PLAPSEGASDVDFDGLGMAKAGLQSLPTHDSPLQRYSEDEPTVP.PSESTDYVAPLTC 1139
Db      1074 PLAPSEGASDVDFDGLGMAKAGLQSLPTHDSPLQRYSEDEPTVP.PSESTDYVAPLTC 1133
Qy      1140 SPOPEYVNVQPVVRQPPSPRPGP.PAARPAAGATLERKTLSPGKGVVXQVPAFGAVEN 1199
Db      1134 SPOPEYVNVQPVVRQPPSPRPGP.PAARPAAGATLERKTLSPGKGVVXQVPAFGAVEN 1193
Qy      1200 PEYLTPOGGAPOHPHPPAPSPADNLYYWDOPPEPAGPSTFKGPTAENPEYLGIDV 1259
Db      1194 PEYLTPOGGAPOHPHPPAPSPADNLYYWDOPPEPAGPSTFKGPTAENPEYLGIDV 1253
Qy      1260 PV 1261
Db      1254 PV 1255

RESULT 8
AAy84780
ID      AAy84780 standard; Protein; 1255 AA.
AC      AAy84780;
DT      08-AUG-2000 (first entry)
XX      Amino acid sequence of the SPLICE erB-2 receptor protein.
XX      SPLICE erB-2 receptor protein; cell transformation disorder; cancer;
XX      tumor cell proliferation; tissue degeneration; arthropathy;
XX      bone resorption; inflammatory disease; degenerative disorder;
XX      wound healing.
OS      Homo sapiens.
PN      M0200020579-A1.
PD      13-APR-2000.
XX      01-OCT-1999; 99MO-CA00912.
XX      02-OCT-1999; 98US-0165192.
XX

```

```

PA      (UYMC-) UNIV MCMASTER.
PI      Muller WJ, Siegel PM;
XX      WPI. 2000-303768/26.
XX      N-PSDB: AAA14812.
XX      Nucleic acid encoding an erB 2 receptor protein designated SPLICE
PT      erB-2, inhibitors of the protein are useful for treatment of cancer -
XX      Claim 3; Fig 2; 60pp; English.
XX
CC      The present sequence represents a SPLICE erB-2 receptor protein. The
CC      protein has an in-frame deletion of 16 amino acids, 2 of which are
CC      conserved cysteine residues, compared to the unspliced protein. The
CC      erB-2 polynucleotide is used to construct probes for detecting
CC      disorders of cell transformation such as cancer. Antibodies to the
CC      protein may be used to detect SPLICE erB-2 in a sample. Agents
CC      (e.g. antisense oligonucleotides) which inhibit the expression of
CC      SPLICE erB-2 are useful for reducing tumor cell proliferation and
CC      treating cancer. Substances which stimulate SPLICE erB-2 are useful
CC      for treating conditions involving damaged cells including conditions
CC      in which degeneration of tissue occurs, such as arthropathy, bone
CC      resorption, inflammatory diseases, degenerative disorders of the
CC      central nervous system and wound healing.
XX
SQ      Sequence 1255 AA:
Query Match      98.2%; Score 6730; DB 21; Length 1255;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 1242; Conservative 8; Mismatches 4; Indels 8; Gaps 2
Qy      1 MELALCRWGLLALLPPGAASFOVCTGDMKRLPASBETHLDMRLHYOGQVQGNL 60
Db      1 MELALCRWGLLALLPPGAASFOVCTGDMKRLPASBETHLDMRLHYOGQVQGNL 60
Qy      61 ELTYLPTNASLSFLODIOEVGVVLAHNOVROVQRLRIYRGTLFFEDNALAVLDNG 120
Db      61 ELTYLPTNASLSFLODIOEVGVVLAHNOVROVQRLRIYRGTLFFEDNALAVLDNG 120
Qy      121 DPLNNTTPVTGASPGRLRELQLSLTEILKGVLLIORNPOLCYQDTILMKDIFHKANQLA 180
Db      121 DPLNNTTPVTGASPGRLRELQLSLTEILKGVLLIORNPOLCYQDTILMKDIFHKANQLA 180
Qy      121 DPLNNTTPVTGASPGRLRELQLSLTEILKGVLLIORNPOLCYQDTILMKDIFHKANQLA 180
Db      121 DPLNNTTPVTGASPGRLRELQLSLTEILKGVLLIORNPOLCYQDTILMKDIFHKANQLA 180
Qy      181 LTLLDITRRAACHPCSPMCKSGKSGESSEDCSLRTYVAGGACRCKPLPTDCHEQC 240
Db      181 LTLLDITRRAACHPCSPMCKSGKSGESSEDCSLRTYVAGGACRCKPLPTDCHEQC 240
Qy      241 AAGCTGKHSDCLACHFNHSGICEJHCPALVYNTDFESMPNPGRYTPASCVTACP 300
Db      241 AAGCTGKHSDCLACHFNHSGICEJHCPALVYNTDFESMPNPGRYTPASCVTACP 300
Qy      301 YNYLSTDVSGCTLVCPJLHNOEVTAEADGTORCEKSNPCARVCYGLGMFNNFTVSEMLR-V 359
Db      301 YNYLSTDVSGCTLVCPJLHNOEVTAEADGTORCEKSNPCARVCYGLGMFNNFTVSEMLR-V 353
Qy      360 PKYSASHLEFFACCKXIFGSLAFLEPSPGDPASNAPIQFPOLOVFETLEITGYLYIS 419
Db      354 RAYTSANIQIFACCKXIFGSLAFLEPSPGDPASNAPIQFPOLOVFETLEITGYLYIS 413
Qy      420 AMPDSLJDSLVFQNTQVIRGRLIHNGAVSLTQIGISWLGRLSRLQSGALATHNHTH 479
Db      414 AMPDSLJDSLVFQNTQVIRGRLIHNGAVSLTQIGISWLGRLSRLQSGALATHNHTH 473
Qy      480 LCFVHTVPMDDQLFRNHQALHTTANRPEDECVCEGLAQHLCARGHCMGPGFTQCVCNSQ 539
Db      474 LCFVHTVPMDDQLFRNHQALHTTANRPEDECVCEGLAQHLCARGHCMGPGFTQCVCNSQ 533
Qy      540 FLRGQCEVECEKYLQGLPREYVNAHCHLPCHPECCPQNGSVTCFPEADQCVACHAYDP 599
Db      534 FLRGQCEVECEKYLQGLPREYVNAHCHLPCHPECCPQNGSVTCFPEADQCVACHAYDP 593
Qy      600 PFCVACRCSGVKPDLSIYPIKPFDEBEGACOPCLNCTHSCVLDLDDGCAEERASPLTS 659

```

```

Db      594 PFCVAPCPSPGKEDLSYMPIMKFPDEGACQCPINCTHSQVLDLKGCPACQSPPLTS 653
Qy      660 IYSAVVGILLVWLVGVVFGILIRKROQKIRKYMRLLOETELVEBLTPSGAMPQAOQR 719
Db      654 IISAVVGLLVVGVVFGILIRKROQKIRKYMRLLOETELVEBLTPSGAMPQAOQR 713
Qy      720 ILKETELRKVKVLSGAGFGTVYKGIWIPDGENYKIPVAIKVLENTSPANKELIDEAVY 779
Db      714 ILKETELRKVKVLSGAGFGTVYKGIWIPDGENYKIPVAIKVLENTSPANKELIDEAVY 773
Qy      760 MAGVSPVYSRLIGICTSTVOLTOLMPYGCCLLDHRENRGSLGSDLLNMCQIAKQM 839
Db      774 MAGVSPVYSRLIGICTSTVOLTOLMPYGCCLLDHRENRGSLGSDLLNMCQIAKQM 833
Qy      840 SYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGKVPKMALES 899
Db      834 SYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGKVPKMALES 893
Qy      900 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGRLLPOPICTIDVYM 959
Db      894 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGRLLPOPICTIDVYM 953
Qy      960 IMVKCMIDSECRPFRELVSFSSRMARDPQRFVYIQNEDLGASPLDSTFYRSLLEDD 1019
Db      954 IMVKCMIDSECRPFRELVSFSSRMARDPQRFVYIQNEDLGASPLDSTFYRSLLEDD 1013
Qy      1020 MGDLYDAEEYLVPOQGFCDPPAPAGAGMHHRRSSSTSGGDDLTLGLEPSEBAPPS 1079
Db      1014 MGDLYDAEEYLVPOQGFCDPPAPAGAGMHHRRSSSTSGGDDLTLGLEPSEBAPPS 1073
Qy      1080 PLAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLORYSEDPVLPSETDGYVAPLTC 1139
Db      1074 PLAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLORYSEDPVLPSETDGYVAPLTC 1133
Qy      1140 SPQPEYVNPQPVRRPQPSREGEPLPAARPAATLERAKTISPGKNGVYKVFAGAVEN 1199
Db      1134 SPQPEYVNPQPVRRPQPSREGEPLPAARPAATLERAKTISPGKNGVYKVFAGAVEN 1193
Qy      1200 PEYLPQCGAARQPPPAFSPAFDNLVYWDQDPERGAPSTFKGTPTAENPEYLGIDV 1259
Db      1194 PEYLPQCGAARQPPPAFSPAFDNLVYWDQDPERGAPSTFKGTPTAENPEYLGIDV 1253
Qy      1260 PV 1261
Db      1254 PV 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
XX
AC AAB85458;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
KW Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KM oncogene; cancer; cytosolic; vaccine; p185; c-erbB2.
XX
OS Homo sapiens.
XX
PN MO200151463-A2.
XX
PD 26-JUL-2001.
XX
PF 19-JAN-2001; 2001WO-US01850.
XX
PR 21-JAN-2000; 2000US-0177545.
XX
PA (CORI-) CORIYA CORP.
XX

```

```

PI Cheever MA, Hand-Zimmermann S;
XX WPI, 2001-476112/51.
DR N-PSDB; AAB23392.
XX
PT New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
PT for treating or preventing cancer, e.g. breast cancer
XX
PS Claim 2, Page 41-46; 49pp; English.
XX
CC The invention provides an isolated antigen-presenting cell, which
CC expresses at least an immunogenic portion of a polypeptide that produces
CC an immune response to HER-2/neu protein. The antigen-presenting cells are
CC useful as vaccines for eliciting or enhancing an immune response to
CC HER-2/neu protein, particularly in treating or preventing malignancies in
CC which the HER-2/neu oncogene is associated. Specifically, these are
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC colon, lung or prostate cancers. The present sequence represents
CC the human HER-2/neu protein (also known as p.185 or c-erbB2).
XX
SQ Sequence 1255 AA;
Query Match 98.2%; Score 6730; DB 22; Length 1255;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 1242; Conservative 8; Mismatches 4; Indels 8; Gaps 2
Qy 1 MELALCRWGLLALALPPGAASQVCTGDMKRLPASPEITHLDMRLHYOGCVVQGNL 60
Db 1 MELALCRWGLLALALPPGAASQVCTGDMKRLPASPEITHLDMRLHYOGCVVQGNL 60
Qy 61 ELTYLPTNASISFLQDIQEVQGVYVLIHANOQVQPIQRLRIYVGTOLFEDNVALVLDNG 120
Db 61 ELTYLPTNASISFLQDIQEVQGVYVLIHANOQVQPIQRLRIYVGTOLFEDNVALVLDNG 120
Qy 121 DPLNNTPTVYGASGGGRELQSLNELKGVLIQRPOLCYQDTITLWMDIFKKNQOLA 180
Db 121 DPLNNTPTVYGASGGGRELQSLNELKGVLIQRPOLCYQDTITLWMDIFKKNQOLA 180
Qy 121 DPLNNTPTVYGASGGGRELQSLNELKGVLIQRPOLCYQDTITLWMDIFKKNQOLA 180
Db 121 DPLNNTPTVYGASGGGRELQSLNELKGVLIQRPOLCYQDTITLWMDIFKKNQOLA 180
Qy 181 LTLIDNRSRACHPCSPMKSGSRGSSSEDQSLRTVYVAGGACARCKGPLPDCHEQC 240
Db 181 LTLIDNRSRACHPCSPMKSGSRGSSSEDQSLRTVYVAGGACARCKGPLPDCHEQC 240
Qy 241 AAGCTGPRHSDCLACIPNHSICELHCPALVYNTDTEFSMNPBGRYFEGASCTYAC 300
Db 241 AAGCTGPRHSDCLACIPNHSICELHCPALVYNTDTEFSMNPBGRYFEGASCTYAC 300
Qy 301 YNYLSTDVGSCTLVCPHNOEVTAEQDTCRCEKSKPCARVYCYGLGMFNFTVSFWLR-V 359
Db 301 YNYLSTDVGSCTLVCPHNOEVTAEQDTCRCEKSKPCARVYCYGLGMFNFTVSFWLR-V 359
Qy 360 PKVSAHLSEEPAGCKKIFGSLAFPEPFDDPASNTAPLPQEQLOVETLEETIGVLYIS 419
Db 360 PKVSAHLSEEPAGCKKIFGSLAFPEPFDDPASNTAPLPQEQLOVETLEETIGVLYIS 419
Qy 354 RAVTSANIQEFAGCKKIFGSLAFPEPFDDPASNTAPLPQEQLOVETLEETIGVLYIS 413
Db 354 RAVTSANIQEFAGCKKIFGSLAFPEPFDDPASNTAPLPQEQLOVETLEETIGVLYIS 413
Qy 420 AMPDLPPLSVFQNLQVIRGRILNNGAYSLTIQSLGSLWGLSLRELISGLALIHNTN 479
Db 420 AMPDLPPLSVFQNLQVIRGRILNNGAYSLTIQSLGSLWGLSLRELISGLALIHNTN 479
Qy 414 AKPDLPLSVFQNLQVIRGRILNNGAYSLTIQSLGSLWGLSLRELISGLALIHNTN 473
Db 414 AKPDLPLSVFQNLQVIRGRILNNGAYSLTIQSLGSLWGLSLRELISGLALIHNTN 473
Qy 480 LCFVHTVPMQDLFNNPHQALHTANREDECVGGLACHQLCARGHWGQGFQCVNCSQ 539
Db 480 LCFVHTVPMQDLFNNPHQALHTANREDECVGGLACHQLCARGHWGQGFQCVNCSQ 539
Qy 474 LCFVHTVPMQDLFNNPHQALHTANREDECVGGLACHQLCARGHWGQGFQCVNCSQ 533
Db 474 LCFVHTVPMQDLFNNPHQALHTANREDECVGGLACHQLCARGHWGQGFQCVNCSQ 533
Qy 540 FLRQGEVCEECRVLOGLPREYVNRHCLPCHEPCQPONGSVTCGFPADQVCAHYKDP 599
Db 540 FLRQGEVCEECRVLOGLPREYVNRHCLPCHEPCQPONGSVTCGFPADQVCAHYKDP 599
Qy 534 FLRQGEVCEECRVLOGLPREYVNRHCLPCHEPCQPONGSVTCGFPADQVCAHYKDP 593
Db 534 FLRQGEVCEECRVLOGLPREYVNRHCLPCHEPCQPONGSVTCGFPADQVCAHYKDP 593
Qy 600 PFCVAPCPSPGKEDLSYMPIMKFPDEGACQCPINCTHSQVLDLKGCPACQSPPLTS 655
Db 594 PFCVAPCPSPGKEDLSYMPIMKFPDEGACQCPINCTHSQVLDLKGCPACQSPPLTS 653
Qy 660 IYSAVVGILLVWLVGVVFGILIRKROQKIRKYMRLLOETELVEBLTPSGAMPQAOQR 715
Db 654 IYSAVVGILLVWLVGVVFGILIRKROQKIRKYMRLLOETELVEBLTPSGAMPQAOQR 713

```

Db 654 IISAVGILLVVLGVVIGIILIKRQOKIRKXKMRLLQETELVEPLTPSGAMENQOMR 713
 QY 720 IKETELRKVKYLGSGAGTGYKGIWIPDGENVKIPVAIKYLRENTSPKANKEILDPAVY 779
 Db 714 IKETELRKVKYLGSGAGTGYKGIWIPDGENVKIPVAIKYLRENTSPKANKEILDPAVY 773
 QY 780 MAGVSPYVSRLLGLICLSTVOLVLTQMPYGGCLDHFENRGRIGSODLLMCMQIAKGM 839
 Db 774 MAGVSPYVSRLLGLICLSTVOLVLTQMPYGGCLDHFENRGRIGSODLLMCMQIAKGM 833
 QY 840 SYLEDVRLVHRDLAARNLVKSPNHVKITDGLARLIDIDETEVHADGGKVPIMWMALES 899
 Db 834 SYLEDVRLVHRDLAARNLVKSPNHVKITDGLARLIDIDETEVHADGGKVPIMWMALES 893
 QY 900 IIRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVM 959
 Db 894 IIRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVM 953
 QY 960 IMVKCMIDSECRPRPRELVSFSPMARDPQRFVITQMEDIGPASPLDSTTYRSLLEDD 1019
 Db 954 IMVKCMIDSECRPRPRELVSFSPMARDPQRFVITQMEDIGPASPLDSTTYRSLLEDD 1013
 QY 1020 MGDLYDAEYLVPOQGFPCPDPAFGAGMHHRRSSSTRSGGDLTLGLPSESEAPRS 1079
 Db 1014 MGDLYDAEYLVPOQGFPCPDPAFGAGMHHRRSSSTRSGGDLTLGLPSESEAPRS 1073
 QY 1080 PLAPSEGAGSDVFDGDLGMAAKGLQSLPTDPSFLQYSEDPTVPLPSETDGYVADLTC 1139
 Db 1074 PLAPSEGAGSDVFDGDLGMAAKGLQSLPTDPSFLQYSEDPTVPLPSETDGYVADLTC 1133
 QY 1140 SPQPEYVNOPDVRPQPSPREGSLPAASPAATLERAKTISPGKNGVYKQVFAFGAVEN 1199
 Db 1134 SPQPEYVNOPDVRPQPSPREGSLPAASPAATLERAKTISPGKNGVYKQVFAFGAVEN 1193
 QY 1200 PEYLTPOGGAPOHPPPAFSPAFDNLVYWDQDPERGAPSTKGTPTAENPEYGLDV 1259
 Db 1194 PEYLTPOGGAPOHPPPAFSPAFDNLVYWDQDPERGAPSTKGTPTAENPEYGLDV 1253
 QY 1260 PV 1261
 Db 1254 PV 1255
 RESULT 10
 AAG88267
 ID AAG88267 standard; Protein; 1255 AA.
 AC AAG88267;
 XX
 DT 11-SEP-2001 (first entry)
 XX
 DE HER2/neu amino acid sequence.
 XX
 KM Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
 KM Immune response; vaccine; tumour; cancer; cytotoxic; immunostimulant;
 KM tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
 OS Homo sapiens.
 XX
 PN WO200141787-A1.
 PD 14-JUN-2001.
 PF 11-DEC-2000; 2000WO-US31591.
 XX
 PR 10-DEC-1999; 99US-0458299.
 XX
 PA (EPIM-) EPIMKUNE INC.
 XX Pikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
 PI Keogh E;
 XX WPI, 2001-374995/39.
 DR

XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing
 PT cellular immune responses for the prevention and treatment of cancer -
 XX
 PS Disclosure, Page 15, 199pp, English.

CC The present invention describes isolated prepared HER2/neu epitopes (1).
 CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
 CC culture in vitro and binds to a complex of an epitope (i), bound to a
 CC human leukocyte antigen (HLA) molecule; (2) a peptide (ii) comprising (i
 CC and a second epitope and the peptide is less than 50 contiguous amino
 CC acids that have 100% identity with a native peptide sequence of HER2/neu
 CC (3) a vaccine composition (iii) comprising (ii) and a pharmaceutical
 CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
 CC (i); and (5) an isolated nucleic acid encoding (ii). (i) has cytotoxic
 CC and immunostimulant activities, and can be used in vaccines. (i), (ii)
 CC and (iii) are useful for inducing cellular immune responses for the
 CC prevention and treatment of cancer. (i) and (ii) are useful for
 CC monitoring or evaluating an immune response to a tumour-associated
 CC antigen when incubated with a T lymphocyte sample from a patient and
 CC detecting the presence of bound T lymphocyte to (i) or (ii). Epitope
 CC based vaccines mean that immunosuppressive epitopes that may be present
 CC in whole antigens may be avoided. Selected epitopes may be combined to
 CC enhance immunogenicity. The possible pathological side effects caused by
 CC infectious agents or whole protein antigen is eliminated. The vaccine
 CC provides the ability to direct and focus an immune response to multiple
 CC selected antigens from the same pathogen. Epitope-based anti-tumour
 CC vaccines provides the opportunity to combine epitopes derived from
 CC multiple tumour-associated molecules addressing the problem of tumour-
 CC tumour variability and reducing the likelihood of tumour escape due to
 CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
 CC the exemplification of the present invention.
 XX

Sequence 1255 AA;

Query Match 98.2%; Score 6730; DB 22; Length 1255;
 Best Local Similarity 98.4%; Pred. No. 0;
 Matches 1242; Conservative 8; Mismatches 4; Indels 8; Gaps :

QY 1 MELALCRWGLLALLPPGAASVCTGTDMKRLPASPETHLDMLHLYOGCQVVOGNT 60
 1 MELALCRWGLLALLPPGAASVCTGTDMKRLPASPETHLDMLHLYOGCQVVOGNT 60
 QY 61 ELYTPTASLSFLDIOEVGVYLIANOVQVPLQRLRIVRGTQVLFEDNYALAVLDNG 120
 61 ELYTPTASLSFLDIOEVGVYLIANOVQVPLQRLRIVRGTQVLFEDNYALAVLDNG 120
 Db 61 ELYTPTASLSFLDIOEVGVYLIANOVQVPLQRLRIVRGTQVLFEDNYALAVLDNG 120
 QY 121 DPLNNTTAVTGASPGGLRETLQRLSELTELKGVYIQRNPOLCYODTTLMKDIFRKNOLA 180
 121 DPLNNTTAVTGASPGGLRETLQRLSELTELKGVYIQRNPOLCYODTTLMKDIFRKNOLA 180
 Db 121 DPLNNTTAVTGASPGGLRETLQRLSELTELKGVYIQRNPOLCYODTTLMKDIFRKNOLA 180
 QY 181 LTLIDTNRSPACHPCSPWCKSGSRGSESDQSLRTVACAGGACRCKGPLPTDCHEQC 240
 181 LTLIDTNRSPACHPCSPWCKSGSRGSESDQSLRTVACAGGACRCKGPLPTDCHEQC 240
 Db 181 LTLIDTNRSPACHPCSPWCKSGSRGSESDQSLRTVACAGGACRCKGPLPTDCHEQC 240
 QY 241 AAGCGPAGSCLCLFHNNHSGICELCPALVYNTDFTFSMPBERRYFGASCYVACP 300
 241 AAGCGPAGSCLCLFHNNHSGICELCPALVYNTDFTFSMPBERRYFGASCYVACP 300
 Db 241 AAGCGPAGSCLCLFHNNHSGICELCPALVYNTDFTFSMPBERRYFGASCYVACP 300
 QY 301 YNYLSTDVGSCTLYCPJHNDCEVTAEDGTORCEKSKPCARVCYGLGMFNFTVSMFLR-V 359
 301 YNYLSTDVGSCTLYCPJHNDCEVTAEDGTORCEKSKPCARVCYGLGMFNFTVSMFLR-V 359
 Db 301 YNYLSTDVGSCTLYCPJHNDCEVTAEDGTORCEKSKPCARVCYGLGMFNFTVSMFLR-V 359
 QY 360 PKVASHLEEFAGCKKIFGSLAFIPSEFDGDPASNTAPLPQLOVFTLEETITGLYIS 419
 360 PKVASHLEEFAGCKKIFGSLAFIPSEFDGDPASNTAPLPQLOVFTLEETITGLYIS 419
 Db 360 PKVASHLEEFAGCKKIFGSLAFIPSEFDGDPASNTAPLPQLOVFTLEETITGLYIS 419
 QY 354 RAVTSANTQEFAGCKKIFGSLAFIPSEFDGDPASNTAPLPQLOVFTLEETITGLYIS 413
 354 RAVTSANTQEFAGCKKIFGSLAFIPSEFDGDPASNTAPLPQLOVFTLEETITGLYIS 413
 Db 354 RAVTSANTQEFAGCKKIFGSLAFIPSEFDGDPASNTAPLPQLOVFTLEETITGLYIS 413
 QY 420 AMPDLPDLAVFYQNLQVIRGRIILNNGAVSLTLOGLGSMGLSLRELGGALLIHNNH 479
 420 AMPDLPDLAVFYQNLQVIRGRIILNNGAVSLTLOGLGSMGLSLRELGGALLIHNNH 479
 Db 414 AMPDLPDLAVFYQNLQVIRGRIILNNGAVSLTLOGLGSMGLSLRELGGALLIHNNH 473
 414 AMPDLPDLAVFYQNLQVIRGRIILNNGAVSLTLOGLGSMGLSLRELGGALLIHNNH 473
 QY 480 LCFVHTVPMQDLFNPQALLHTANRDEDECVBGLACQGLCARGCWGPPTQVNCQ 539
 480 LCFVHTVPMQDLFNPQALLHTANRDEDECVBGLACQGLCARGCWGPPTQVNCQ 539


```

Db      474 LCFVHTVPMDOQLFRNPHQALLHTANRPEDCEVSGGLACHQLCARGHGMGPPTQCVNCSQ 533
Qy      540 FLRGOCEVEBCRVLQGLPREYVNAHCLPCHEPCQPNQNGSVTCFGBPADQVCAAHYKDP 599
Db      534 FLRGOCEVEBCRVLQGLPREYVNAHCLPCHEPCQPNQNGSVTCFGBPADQVCAAHYKDP 593
Qy      600 PFCVAPRCSSGVKPDLSVYMKPKFPDEGACOPCINTCTHSCVLLDDKGCAGAEASPLTS 659
Db      594 PFCVAPRCSSGVKPDLSVYMKPKFPDEGACOPCINTCTHSCVLLDDKGCAGAEASPLTS 653
Qy      660 IVSAVAVGILLVVLGVVFGILLIKRQOKIRKRYTMRLLQETELVEPLTPSGAMPNOAKR 719
Db      654 IISAVVGLLVVLGVVFGILLIKRQOKIRKRYTMRLLQETELVEPLTPSGAMPNOAKR 713
Qy      720 IUKETELKRVKVLGSGAGTGYKGIWIPDGNVYKIPVAIKVLRNTSPKANKILLDEAY 779
Db      714 IUKETELKRVKVLGSGAGTGYKGIWIPDGNVYKIPVAIKVLRNTSPKANKILLDEAY 773
Qy      780 MAGVSPYVSRLLGICLTSTVQLVTQLMPIYGCILLDHVRENGRLGSODLLNMCQIAKGM 839
Db      774 MAGVSPYVSRLLGICLTSTVQLVTQLMPIYGCILLDHVRENGRLGSODLLNMCQIAKGM 833
Qy      840 SYLEDVRLVHRDLAARNVLYVSPNVKITTDFGLARLLDIDETEHADGCKVPIKMALES 899
Db      834 SYLEDVRLVHRDLAARNVLYVSPNVKITTDFGLARLLDIDETEHADGCKVPIKMALES 893
Qy      900 IIRRRFTHOSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYM 959
Db      894 IIRRRFTHOSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYM 953
Qy      960 IMVKCMWIDSECRPFRELIVSEFSMAADPORFVYIQNEDIGPASPILDSTFYRSLLDDO 1019
Db      954 IMVKCMWIDSECRPFRELIVSEFSMAADPORFVYIQNEDIGPASPILDSTFYRSLLDDO 1013
Qy      1020 MGDIYDAEEYLVPOGCFPCPDPAAGAMVHRRSSSTRSGGDLTLGEPSEEPAPRS 1079
Db      1014 MGDIYDAEEYLVPOGCFPCPDPAAGAMVHRRSSSTRSGGDLTLGEPSEEPAPRS 1073
Qy      1080 PLAPSEGASDVFDGDLGMAAKGLQSLPTHDPSPLOQYSEDPVPLPSETDGYVAPLTC 1139
Db      1074 PLAPSEGASDVFDGDLGMAAKGLQSLPTHDPSPLOQYSEDPVPLPSETDGYVAPLTC 1133
Qy      1140 SPQPEYVNPDPVPRQPPSPREGLPLAARPAAGATLERAKTSLSPKNGVYKDVAFAGAVEN 1199
Db      1134 SPQPEYVNPDPVPRQPPSPREGLPLAARPAAGATLERAKTSLSPKNGVYKDVAFAGAVEN 1193
Qy      1200 PEYLTPQGGAAPQHPPPAFSPAFNLYYWDQDPBERGAPSTFKGPTJANPEYLGIDV 1259
Db      1194 PEYLTPQGGAAPQHPPPAFSPAFNLYYWDQDPBERGAPSTFKGPTJANPEYLGIDV 1253
Qy      1260 PV 1261
Db      1254 PV 1255

```

RESULT 11

AAE24067 standard; Protein; 1255 AA.

```

ID      AAE24067
AC      AAE24067;
DT      23-SEP-2002 (first entry)
DE      Human Her-2 protein.
KW      Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW      hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW      tumour; gene therapy; phosphothioate backbone.
OS      Homo sapiens.
PN      MO200222636-A1.

```

```

PD      21-MAR-2002.
XX      XX
XX      PF 12-SEP-2001; 2001MO-US28572.
XX      PR 15-SEP-2000; 2000US-0663834.
XX      PA (ISIS-) ISIS PHARM INC.
XX      PI Bennett CF, Cowsett LM;
XX      XX
XX      DR WPI; 2002-471192/50.
XX      DR N-PsDB; AAD38904.
XX      PT Novel antisense oligonucleotide which modulates the expression of Human
XX      PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
XX      PT inflammation or to prevent infection in humans -
XX      TS Example 13; Page 95-107; 116pp; English.
XX      CC The invention relates to antisense compounds targeted to a nucleic
XX      CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
XX      CC that specifically hybridizes with and inhibits the expression of Her2.
XX      CC Antisense compounds of the invention are used for treating diseases or
XX      CC conditions associated with Her2 such as hyperproliferative disorders
XX      CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
XX      CC neural or cardiac cancer. They are also useful prophylactically e.g.
XX      CC to prevent or delay infection, inflammation and tumour formation. The
XX      CC invention is also used in gene therapy. The present sequence is human
XX      CC Her2 protein.
XX      SQ Sequence 1255 AA;
XX      XX
XX      Query Match 98.2%; Score 6730; DB 23; Length 1255;
XX      Best Local Similarity 98.4%; Pred. No. 0;
XX      Matches 1242; Conservative 8; Mismatches 4; Indels 8; Gaps 2

```



```

Db      534 FLRGQECVEECRYLQGIIPREYVNAHCLPCHPECCQPNQSVICFGEADQCAVCAHYXDP 593
Qy      600 PFCVARGSPGVKEDLSYMPIMKFPDEGACQCPICINCTHSCVDLDKXGPAORASPLTS 659
Db      594 PFCVARGSPGVKEDLSYMPIMKFPDEGACQCPICINCTHSCVDLDKXGPAORASPLTS 653
Qy      660 IVSAVVGILLVVLGVVFGILLIKRQOKIRKYTMRLLOETELVEELTPSGAMPQAOQR 719
Db      654 IISAVVGLLVVLGVVFGILLIKRQOKIRKYTMRLLOETELVEELTPSGAMPQAOQR 713
Qy      720 ILKETEELRKRVUGSGAFCTVYKGIWIPDGENKIVAIKVLRNNSPANKELIDEAVY 779
Db      714 ILKETEELRKRVUGSGAFCTVYKGIWIPDGENKIVAIKVLRNNSPANKELIDEAVY 773
Qy      780 MAGVSPYVSRLLGICLTSTVQLVTCIMPYGCLLDHVRNKGHLSGQDLINMCQIJAQM 839
Db      774 MAGVSPYVSRLLGICLTSTVQLVTCIMPYGCLLDHVRNKGHLSGQDLINMCQIJAQM 833
Qy      840 SYLEDVFLVHDLAARVNLVKSNNHVKITDFGLARLLDIDETEHADGKVPDKMALES 899
Db      834 SYLEDVFLVHDLAARVNLVKSNNHVKITDFGLARLLDIDETEHADGKVPDKMALES 893
Qy      900 ILRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYM 959
Db      894 ILRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYM 953
Qy      960 IMVKCMTIDSECPREFELVSEFSRMAADPQRFVVIQNEIDLGPASPLDSTFYRSLLDD 1019
Db      954 IMVKCMTIDSECPREFELVSEFSRMAADPQRFVVIQNEIDLGPASPLDSTFYRSLLDD 1013
Qy      1020 MGDVDAEEYLVPOQGFCEPDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEPARS 1079
Db      1014 MGDVDAEEYLVPOQGFCEPDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEPARS 1073
Qy      1080 PLAPSEGAGSDVFDGDLGMAANGLOSLPTHDPSPLOQYSEDPVLPBETDGYVAPLTC 1139
Db      1074 PLAPSEGAGSDVFDGDLGMAANGLOSLPTHDPSPLOQYSEDPVLPBETDGYVAPLTC 1133
Qy      1140 SPOEYVNOQDVAPQPSRREGPLPAARPGATLLEBAKTLSPKXGVYCDVAFGAGVEN 1199
Db      1134 SPOEYVNOQDVAPQPSRREGPLPAARPGATLLEBAKTLSPKXGVYCDVAFGAGVEN 1193
Qy      1200 PEYLTPOGGAAPQHPAPFAPFNDLYWDQPPERGAAPPSTFKGTPTAENPEYLGIDV 1259
Db      1194 PEYLTPOGGAAPQHPAPFAPFNDLYWDQPPERGAAPPSTFKGTPTAENPEYLGIDV 1253
Qy      1260 PV 1261
Db      1254 PV 1255

RESULT 12
AAE20479 standard; Protein; 1255 AA.
ID AAE20479;
XX AAE20479;
AC AAE20479;
AAE20479;
DT 01-JUL-2002 (first entry)
XX 01-JUL-2002 (first entry)
DE Human Her-2/neu protein.
XX Human Her-2/neu protein.
XX Human Her-2/neu protein; immune response; gene therapy; breast cancer;
XX human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX Homo sapiens.
XX Homo sapiens.
FH Key Location/Qualifiers
FT Region 1021..1030
XX /note="Naturally processed HLA-B44-restricted epitope"
PN W0200214503-A2.
XX

```

```

PD 21-FEB-2002.
XX
XX 14-AUG-2001; 2001WO-US61733.
PF 14-AUG-2001; 2001WO-US61733.
XX 14-AUG-2000; 2000US-225152P.
XX 28-SEP-2000; 2000US-226428P.
FR 21-FEB-2001; 2001US-270520P.
XX
XX (CORI-) CORIYA CORP.
PA
XX Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
PI McNeill PD, Vedvick TS;
XX WPI; 2002-280758/32.
DR N-PSDB; AAD32743.
XX
XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT prevention and diagnosis of cancer, preferably breast cancer
XX
XX Disclosure; Page 114-117; 129pp; English.
PS
XX The invention relates to an isolated Her-2/Neu polypeptide composition
XX effective for eliciting an immune response. The invention is useful for
XX eliciting an immune response in a patient, where the patient is human
XX leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
XX The composition is useful for the therapy and diagnosis of cancer,
XX preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
XX and other compositions for the diagnosis, prevention and treatment of
XX human malignancies, for stimulating and/or expanding T cells specific for
XX Her-2/Neu polypeptide and for inhibiting the development of cancer in a
XX patient. The invention is useful for stimulating a T cell response in a
XX human patient, as probe or primer for nucleic acid hybridisation, to
XX selectively form duplex molecules with complementary stretches of the
XX selected Her-2/Neu gene or gene fragments of interest, to isolate a full
XX length gene from a suitable library, and to direct expression of a
XX polypeptide in appropriate host cells. The composition is useful in
XX prophylactic or therapeutic applications and for the treatment of cancer,
XX preferably for the immunotherapy of breast cancer and other Her-2/Neu-
XX associated malignancies. The invention is useful in gene therapy. The
XX present sequence is human Her-2/neu protein.
SQ
Sequence 1255 AA;
Query Match 98.2%; Score 6730; DB 23; Length 1255;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 1242; Conservative 8; Mismatches 4; Indels 8; Gaps 2
Qy 1 MELDALCRWGLLALLPFGAASGVCTGDMKRLPASPETHLDMRLHYGSCVVGNTL 60
Db 1 MELDALCRWGLLALLPFGAASGVCTGDMKRLPASPETHLDMRLHYGSCVVGNTL 60
Qy 61 ELTYLPTNASTSLFODIOEVQGVVLLAHNOVQVPLQRLIRVGTQLFEDNALAVLNG 120
Db 61 ELTYLPTNASTSLFODIOEVQGVVLLAHNOVQVPLQRLIRVGTQLFEDNALAVLNG 120
Qy 121 DPLNNTPTVYGASPGGLRELQRLSLTEILKGGVLIQRNQLCYODTILMKDFHKNQCLA 180
Db 121 DPLNNTPTVYGASPGGLRELQRLSLTEILKGGVLIQRNQLCYODTILMKDFHKNQCLA 180
Qy 181 LTLIDNRSRACHPCSPCKSGKSCWESSSEDCOSLTRIVYACAGCARCKGRLPTDCCHQC 240
Db 181 LTLIDNRSRACHPCSPCKSGKSCWESSSEDCOSLTRIVYACAGCARCKGRLPTDCCHQC 240
Qy 241 AAGCTGPKSDCLACJHFNHSGICELHCALVTYNTDFESMPNPEGRTFFASCVTACP 300
Db 241 AAGCTGPKSDCLACJHFNHSGICELHCALVTYNTDFESMPNPEGRTFFASCVTACP 300
Qy 301 YNYLSTDVSGCTLVCPHLNQEVTAEDGTQRCCKSPCARVYCYGLGFENNTVSWFLR-V 359
Db 301 YNYLSTDVSGCTLVCPHLNQEVTAEDGTQRCCKSPCARVYCYGLGFENNTVSWFLR-V 359
Qy 360 PKVSASHLEEFACCKKIPQSLAFLEPSPFGDPSANTAPQAPQLOVFETLEITGYLTS 419
Db 360 PKVSASHLEEFACCKKIPQSLAFLEPSPFGDPSANTAPQAPQLOVFETLEITGYLTS 419

```

```

Db      354  RAVTSAIQEPAGCKITFGSLAFLPESFDGDPASNTAPLQPEQVQVETLEBITGYLTIS 413
QY      420  AMPDSLPLSLVFOULQVIRGRILHNGAYSLTLQQLGSMIGLSLHELGLALIHNNH 479
Db      414  AMPDSLPLSLVFOULQVIRGRILHNGAYSLTLQQLGSMIGLSLHELGLALIHNNH 473
QY      480  LCFVHTVMDQLFERNPHQALHTANRPEDECEVGEGLACHOLCARGHCMWGPPTVCNCSQ 539
Db      474  LCFVHTVMDQLFERNPHQALHTANRPEDECEVGEGLACHOLCARGHCMWGPPTVCNCSQ 533
QY      540  FLRGECEVECEVRVLOGLPREYVNAHCLPCHECECPONSVTGGPPADCCVACAHYKCP 599
Db      534  FLRGECEVECEVRVLOGLPREYVNAHCLPCHECECPONSVTGGPPADCCVACAHYKCP 593
QY      600  PFCVAPCPSPVLPDLSTYMPWKPFDEEGACOPPCINCTHSCVDLDDKGCABEASAPLTS 659
Db      594  PFCVAPCPSPVLPDLSTYMPWKPFDEEGACOPPCINCTHSCVDLDDKGCABEASAPLTS 653
QY      660  IVSAVVGILLVVLGVVFGILIKRQOKIRKXTMRLLQETELVEPLTPSGAMPNQAQR 719
Db      654  IISAVVGILLVVLGVVFGILIKRQOKIRKXTMRLLQETELVEPLTPSGAMPNQAQR 713
QY      720  ILKETELRKVVLGSGAFGYKGIWIPDGENVKIPVAIKVLSENTPSKAKXELDEAVY 779
Db      714  ILKETELRKVVLGSGAFGYKGIWIPDGENVKIPVAIKVLSENTPSKAKXELDEAVY 773
QY      780  MAGVSPVYSRLGLCLTSTVQLVTLQMPYGCLLDHVENRGRGLGSDLLMCMQIAKGM 839
Db      774  MAGVSPVYSRLGLCLTSTVQLVTLQMPYGCLLDHVENRGRGLGSDLLMCMQIAKGM 833
QY      840  SYLEDVRLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGVPIKMMALSS 899
Db      834  SYLEDVRLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGVPIKMMALSS 893
QY      900  ILRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPICTIDVYM 959
Db      894  ILRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPICTIDVYM 953
QY      960  IMVKCMIMDSECRPFRELVSEFSRMAADPQRFVITQNEDLGPAASPLDSTFFYSLLDD 1019
Db      954  IMVKCMIMDSECRPFRELVSEFSRMAADPQRFVITQNEDLGPAASPLDSTFFYSLLDD 1013
QY      1020  MODLVDAEEVLPQOGFPCPPAPGAGGMVHRHRSSTSGGDLTLGLEPSEEEZAPRS 1079
Db      1014  MODLVDAEEVLPQOGFPCPPAPGAGGMVHRHRSSTSGGDLTLGLEPSEEEZAPRS 1073
QY      1080  PLAPSEGAAGSVFDGDLGMAKAGLQSLPTHDPSPICQRYSEDPVPIPSSETDGYVAPLTC 1139
Db      1074  PLAPSEGAAGSVFDGDLGMAKAGLQSLPTHDPSPICQRYSEDPVPIPSSETDGYVAPLTC 1133
QY      1140  SPOPEYVNPDPVRPSPSPREGPIPAARPAAGATLEBAKTLSPGKNGVVKVYFAGGAVEN 1199
Db      1134  SPOPEYVNPDPVRPSPSPREGPIPAARPAAGATLEBAKTLSPGKNGVVKVYFAGGAVEN 1193
QY      1200  PEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPBEGAPSTFKGTPTAENEVYGLDV 1259
Db      1194  PEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPBEGAPSTFKGTPTAENEVYGLDV 1253
QY      1260  PY 1261
Db      1254  PY 1255

```

RESULT 13
AAMS1143
ID AAMS1143 standard; Protein; 1255 AA.

AC AAMS1143;

DT 17-JUN-2002 (first entry)

DE Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX

```

KW Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
KX tyrosine kinase; receptor; c-erbB2; gene therapy.
XX
XX Homo sapiens.
OS
XX
XX Key
FH Location/Qualifiers
FT 1..653
FT Domain
FT /note= "extracellular domain"
FT 676..1255
FT Domain
FT /note= "intracellular domain"
FT 990..1255
FT /note= "phosphorylation domain"
PN
PD WO200212341-A2.
PD 14-FEB-2002.
PD 03-AUG-2001; 2001WO-US24283.
PD 03-AUG-2000; 2000US-0632507.
PD 03-AUG-2000; 2000US-0632507.
PR
XX
XX (CORI-) CORIXA CORP.
XX (SMK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX Cheever MA, Gheysen D;
XX WPI; 2002-241743/29.
XX N-PSDB; ABA92250.
XX
XX Her-2/neu fusion protein for treating or preventing cancer by eliciting
PT or enhancing an immune response to the protein, has Her-2/neu
PT extracellular domain fused to Her-2/neu intracellular or
PT phosphorylation domain
XX
XX Claim 68; Fig 7; 14pp; English.
XX
XX The present sequence is that of human Her-2/neu (p185 glycoprotein
CC or c-erbB2), an oncogenic self-protein and target for anti-cancer
CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
CC in a variety of cancers, including breast, ovarian, colon, lung and
CC prostate cancer. Her-2/neu is a member of the tyrosine kinase
CC family of receptor-like glycoproteins. It comprises an extracellular
CC domain with homology to the epidermal growth factor receptor
CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
CC intracellular domain that also shows homology to EGF. Its
CC overexpression correlates with a poor prognosis in breast and
CC ovarian cancers. The invention provides Her-2/neu fusion
CC proteins, nucleic acids encoding them, viral vectors, and vaccines
CC comprising the fusion proteins or nucleic acid molecules. In
CC preferred fusion proteins, the extracellular domain of a Her-2/neu
CC protein is fused to a Her-2/neu intracellular domain or
CC phosphorylation domain (or its DeltaCD fragment). An immune
CC response to Her-2/neu protein is elicited or enhanced by
CC administering the fusion protein in the form of a vaccine, or by
CC transfecting cells of an animal ex vivo with a nucleic acid
CC encoding the fusion protein, and delivering the transfected cells
CC to the animal. The fusion proteins, nucleic acids, and isolated
CC specific T-cells are useful for inhibiting the development of a
CC cancer, especially breast, ovarian, colon, lung or prostate cancer
CC in a patient. T cells that specifically react with a Her-2/neu
CC fusion protein can be used to remove tumour cells from a sample in
CC order to inhibit the development of cancer in a patient.
XX
XX Sequence 1255 AA:
SQ

```

Query Match 98.2%; Score 6730; DB 23; Length 1255;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 1242; Conservative 8; Mismatches 4; Indels 8; Gaps 2

QY 1 MELALCRWGLLALLPPGAASIVCGTGMKRLPASPETHLDMRHLYGCGVQVQGNL 60
DB 1 MELALCRWGLLALLPPGAASIVCGTGMKRLPASPETHLDMRHLYGCGVQVQGNL 60

```

QY 61 ELYLPFNASLFLDIOEVQGYVLIANOVROVPLQRLIRVGTQIFEDNYALAVLDNG 120
DB 61 ELYLPFNASLFLDIOEVQGYVLIANOVROVPLQRLIRVGTQIFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGASPGGLRELOLRSTTELKGGVLIQRNPOLCYDDTLMDIFHKNOLA 180
DB 121 DPLNNTTPTVGASPGGLRELOLRSTTELKGGVLIQRNPOLCYDDTLMDIFHKNOLA 180
QY 181 LTLIDTNRSRACHPCSPMCGRSCWGESSEDCQSLTRTVCAAGCARGKGPJPTDCHEQC 240
DB 181 LTLIDTNRSRACHPCSPMCGRSCWGESSEDCQSLTRTVCAAGCARGKGPJPTDCHEQC 240
QY 241 AAGCTGPKHSDCLAHFNHSGICELHCPALVTYNTDFFEMRPERGRYTFGASCYACAP 300
DB 241 AAGCTGPKHSDCLAHFNHSGICELHCPALVTYNTDFFEMRPERGRYTFGASCYACAP 300
QY 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTQCEKSKPCARCYGLGMFNNFTVSMFLR-V 359
DB 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTQCEKSKPCARCYGLGMFNNFTVSMFLR-V 359
QY 360 PKVSASHLEEFAGCKTIFGSLAFLEPSFDGDPASNTAPLQEQLOVETLEBITGYLIS 419
DB 360 PKVSASHLEEFAGCKTIFGSLAFLEPSFDGDPASNTAPLQEQLOVETLEBITGYLIS 419
QY 420 AMPDSLPLSVFQNLQVIRGRILHNGAVSLTLQGLGISMGLRSLRELGSGLALIHNTH 479
DB 420 AMPDSLPLSVFQNLQVIRGRILHNGAVSLTLQGLGISMGLRSLRELGSGLALIHNTH 479
QY 440 LCFVHTVPMDOLEFNNPHOALLHTANRPDESCVGGGLACHOLCARGHGWGPPTQVNCQ 539
DB 440 LCFVHTVPMDOLEFNNPHOALLHTANRPDESCVGGGLACHOLCARGHGWGPPTQVNCQ 539
QY 540 FLRGQCEVCECRVQGLPREYVNAHCLPCHPEOPONGSTCGEPADQCVACAHYKOP 599
DB 540 FLRGQCEVCECRVQGLPREYVNAHCLPCHPEOPONGSTCGEPADQCVACAHYKOP 599
QY 594 PFCVARCPSGVKPDLSYMPIMKFPDEBGAQCPGICNTHSCVDLDKGCBAEGRASPLTS 659
DB 594 PFCVARCPSGVKPDLSYMPIMKFPDEBGAQCPGICNTHSCVDLDKGCBAEGRASPLTS 659
QY 660 TVSAVVGILLVVVGVVFGILLIKRQOKIRKTYRRLLOETELVEPLTBSGAMPNOQAKR 719
DB 660 TVSAVVGILLVVVGVVFGILLIKRQOKIRKTYRRLLOETELVEPLTBSGAMPNOQAKR 719
QY 720 ILKETELKXKVLGSGAGTVYKGIWIPDGENYKIPAIKYLRENTSPKXKEILLDEAYV 779
DB 720 ILKETELKXKVLGSGAGTVYKGIWIPDGENYKIPAIKYLRENTSPKXKEILLDEAYV 779
QY 780 MAGVSPYVSRLGICLTSTVQVLTQMPYGCCLIDHVRNRRGLGSGDILNMCQIAKGM 839
DB 780 MAGVSPYVSRLGICLTSTVQVLTQMPYGCCLIDHVRNRRGLGSGDILNMCQIAKGM 839
QY 840 SYLEEDVRJNHRDLAARNLVKSPNHVKITPDEGLARLIDIDETVYHNDGGVPIKXWALS 899
DB 840 SYLEEDVRJNHRDLAARNLVKSPNHVKITPDEGLARLIDIDETVYHNDGGVPIKXWALS 899
QY 900 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPPPICTIDVYM 959
DB 900 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPPPICTIDVYM 959
QY 960 IMVKCMWIDSCRFREFELVSEFSRMAPDQRFVYIQLNEDLGASPLDSTFFYSLLEDD 1019
DB 960 IMVKCMWIDSCRFREFELVSEFSRMAPDQRFVYIQLNEDLGASPLDSTFFYSLLEDD 1019
QY 1020 MGDLDVDAEBEYLVPQOGFCDPAPAGAGMYHRRSSSTRSGGADTLTIGLEPSEBAPRS 1079
DB 1020 MGDLDVDAEBEYLVPQOGFCDPAPAGAGMYHRRSSSTRSGGADTLTIGLEPSEBAPRS 1079
QY 1080 PLASEGAGSDVFEGDILGMAKAGLQSLPTHDSPLQRYSEDPTVPLPSETDGYVAPLTC 1139
DB 1080 PLASEGAGSDVFEGDILGMAKAGLQSLPTHDSPLQRYSEDPTVPLPSETDGYVAPLTC 1139
QY 1140 SPOEYVNVQPDVRRQPPSPREGPLPARPAGATLERAKTILSPKNGVAVDVFAGAVEN 1199

```

```

DB 1134 SPOEYVNVQPDVRRQPPSPREGPLPARPAGATLERAKTILSPKNGVAVDVFAGAVEN 1199
QY 1200 PEYLPQGAAPQHPHPPAFSPARDNLYYMQDPEEGAPSPSTKGTPTAENPEYLGLDV 125
DB 1194 PEYLPQGAAPQHPHPPAFSPARDNLYYMQDPEEGAPSPSTKGTPTAENPEYLGLDV 125
QY 1260 PV 1261
DB 1254 PV 1255

RESULT 14
AA077114
ID AA077114 standard; Protein, 1255 AA.
XX
XX AA077114;
XX
DT 05-JUN-2002 (first entry)
XX
XX Human Her-2/neu polypeptide.
DE
XX Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX
XX Homo sapiens.
XX
XX WC020213847-A2.
XX
XX 21-FEB-2002.
XX
XX 13-AUG-2001; 2001WO-US25408.
XX
XX 14-AUG-2000; 2000US-0638280.
XX
XX 28-SEP-2000; 2000US-0675904.
XX
XX (CORI-) CORIYA CORP.
XX
XX Gaiger A, Cheever MA, Hand-zimmermann S;
XX WPI, 2002-280741/32.
XX
XX N-PSDB; ABK10730.
XX
XX Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide
XX
XX Disclosure; Page 71-74; 74pp; English.
XX
XX The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
XX Sequence 1255 AA;
SQ
Query Match 98.2%; Score 6730; DB 23; Length 1255;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 1242; Conservative 8; Mismatches 4; Indels 8; Gaps 2
QY 1 MELALCRWGLLALPPGAASOVCTGIDMKRLPASPETHLMDLRHLYOGQVVGNTL 60
DB 1 MELALCRWGLLALPPGAASOVCTGIDMKRLPASPETHLMDLRHLYOGQVVGNTL 60
QY 61 ELYLPFNASLFLDIOEVQGYVLIANOVROVPLQRLIRVGTQIFEDNYALAVLDNG 120

```

```

Db      61 ELTYLPTNASLSFLDIOEVGYVLIANQVRQVPLQRLIVRGTOLFEDNYALAVLDNG 120
Qy      121 DPLNNTTPVTGASPGGLRELOLRSLTELKGGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
Db      121 DPLNNTTPVTGASPGGLRELOLRSLTELKGGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
Qy      181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACGAGCARCKGPLPTDCCHEQC 240
Db      181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACGAGCARCKGPLPTDCCHEQC 240
Qy      241 AAGCTGPKASDCLACLHFNHSGICELHCPALVYTYTDTFESMPPEGRYTFGASCTVACP 300
Db      241 AAGCTGPKASDCLACLHFNHSGICELHCPALVYTYTDTFESMPPEGRYTFGASCTVACP 300
Qy      301 YNYLSTDVGSCTLVCPLEHNOEVTAEQGTORCEKSKPCARCYGCGMNNFTVAFMLR-V 359
Db      301 YNYLSTDVGSCTLVCPLEHNOEVTAEQGTORCEKSKPCARCYGCGMNNFTVAFMLR-V 359
Qy      360 PRVSASHLEEFAGCKKIFGSLAFLESEFDGPASNTARLOEQLOVETLEITGYLYIS 419
Db      360 PRVSASHLEEFAGCKKIFGSLAFLESEFDGPASNTARLOEQLOVETLEITGYLYIS 419
Qy      420 AMPDSLPLDSVFQNIQVTRGRIHNGAYSILTQGLIGTWLGLRELGSGLAIHHNTH 479
Db      420 AMPDSLPLDSVFQNIQVTRGRIHNGAYSILTQGLIGTWLGLRELGSGLAIHHNTH 479
Qy      479 LCFVHTVPMDQLFRNPHQALHTANRPEDECVEGLACHQLCASGHCMPGPTQCVNCSQ 539
Db      479 LCFVHTVPMDQLFRNPHQALHTANRPEDECVEGLACHQLCASGHCMPGPTQCVNCSQ 539
Qy      540 FLRGOECVEECVLOGLPREYVNAHCLPCHPECPQNGSTTCGPEPDDQVCAAHKDP 599
Db      540 FLRGOECVEECVLOGLPREYVNAHCLPCHPECPQNGSTTCGPEPDDQVCAAHKDP 599
Qy      599 PCVACRCPGKVPDLSTYMPIMKFPDEEGACQPCPNTCHSCVDDDDGCPAECASPLTS 659
Db      599 PCVACRCPGKVPDLSTYMPIMKFPDEEGACQPCPNTCHSCVDDDDGCPAECASPLTS 659
Qy      660 IYSAVVGILLVVLGVVFGILIKRQOKIRKYMRLIQETELVEPLTPSGAMNQOMR 719
Db      660 IYSAVVGILLVVLGVVFGILIKRQOKIRKYMRLIQETELVEPLTPSGAMNQOMR 719
Qy      720 ILKETELRKVKYLGSGAFGTYYKGIWPDGENVKIPVAIKYLRNTPSKAKELIDAVY 779
Db      720 ILKETELRKVKYLGSGAFGTYYKGIWPDGENVKIPVAIKYLRNTPSKAKELIDAVY 779
Qy      780 MAGVGSPIVSRLLIGICTISTVOLVTQLMPEYGLDHYVENSGRIGSODLLMCMQIAKCM 839
Db      780 MAGVGSPIVSRLLIGICTISTVOLVTQLMPEYGLDHYVENSGRIGSODLLMCMQIAKCM 839
Qy      840 SYLEDVRLVHRDLAARNVLVKSNNHVKITDPEGLARLIDITEVYADGCKPIKMALES 899
Db      840 SYLEDVRLVHRDLAARNVLVKSNNHVKITDPEGLARLIDITEVYADGCKPIKMALES 899
Qy      900 ILRRRFTHQSDVMSYGVTVMEITGAKPYDGIIPAREIPDLEKGERLPPPICTIDVYM 959
Db      900 ILRRRFTHQSDVMSYGVTVMEITGAKPYDGIIPAREIPDLEKGERLPPPICTIDVYM 959
Qy      960 IMYKCMWIDSECRPRRRELIVSEFSKMAADPQRFVVIQNEIDIGPASPILDSTRYRLLEDD 1019
Db      960 IMYKCMWIDSECRPRRRELIVSEFSKMAADPQRFVVIQNEIDIGPASPILDSTRYRLLEDD 1019
Qy      1020 MGDVDAEEYLVFOQGFPCPDPAAGAGVHHRHSSSTRSGGDLTLGLEPSEEAAPRS 1079
Db      1020 MGDVDAEEYLVFOQGFPCPDPAAGAGVHHRHSSSTRSGGDLTLGLEPSEEAAPRS 1079
Qy      1080 PLAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSTLQSYSEDPVPLPSEFDGVAAPLTC 1139
Db      1080 PLAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSTLQSYSEDPVPLPSEFDGVAAPLTC 1139
Qy      1140 SPOEYVNOQDVRPOPSPREGPLPAARPAAGATLERAKTLPQKGVGVYDFAFGAVEN 1199

```

```

Db      1134 SPOEYVNOQDVRPOPSPREGPLPAARPAAGATLERAKTLPQKGVGVYDFAFGAVEN 1199
Qy      1200 PEYLTPGQGAAPQHPAPPAPSPAFNDLYYMDQDPPERGAPSTFKGPTAENPEYGLDV 125
Db      1194 PEYLTPGQGAAPQHPAPPAPSPAFNDLYYMDQDPPERGAPSTFKGPTAENPEYGLDV 125
Qy      1260 PV 1261
Db      1254 PV 1255

RESULT 15
AAR39568
ID AAR39568 standard; Protein; 1433 AA.
XX
AC AAR39568;
XX
DT 07-FEB-1994 (first entry)
XX
DE Sequence of c-erbB-2 tumour antigen.
XX
KM Tumour antigen; c-erbB-2; glycoprotein.
XX
OS Homo sapiens.
XX
PN W09316185-A.
XX
PD 19-AUG-1993.
XX
PF 05-FEB-1993; 93MO-US01055.
XX
PR 06-FEB-1992; 92US-0831967.
XX
PA (CETU) CETUS ONCOLOGY CORP.
PA (CREA-) CREATIVE BIOMOLECULES INC.
PI Houston IL, Huston JS, Oppermann H, Ring DB;
DR WPI, 1993-272889/34.
DR N-PSDB; AAQ46083.
XX
PT New single chain Fv polypeptide binding to C-erbB-2 tumour
XX antigen - for imaging or treating breast or ovarian cancer etc.
XX
PS Disclosure; pages 48-54, 87pp; English.
XX
CC c-erbB-2 refers to a protein antigen expressed on the surface of
XX tumour cells, such as breast and ovarian tumour cells, which is an
XX approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
XX pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents
XX the location of a stop codon in AAQ46083.
XX
SQ Sequence 1433 AA;

Query Match 97.6%; Score 6687; DB 14; Length 1433;
Best Local Similarity 97.9%; Pred. No. 0;
Matches 125; Conservative 10; Mismatches 9; Indels 8; Gaps 1;

Qy      1 MEALALCKWGLLALLPFGAASTGYCTDMKLLPASPETHLMLHLLVQGCYVQGNL 60
Db      1 MEALALCKWGLLALLPFGAASTGYCTDMKLLPASPETHLMLHLLVQGCYVQGNL 60
Qy      61 ELTYLPTNASLSFLDIOEVGYVLIANQVRQVPLQRLIVRGTOLFEDNYALAVLDNG 120
Db      61 ELTYLPTNASLSFLDIOEVGYVLIANQVRQVPLQRLIVRGTOLFEDNYALAVLDNG 120
Qy      121 DPLNNTTPVTGASPGGLRELOLRSLTELKGGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
Db      121 DPLNNTTPVTGASPGGLRELOLRSLTELKGGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
Qy      181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACGAGCARCKGPLPTDCCHEQC 240
Db      181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACGAGCARCKGPLPTDCCHEQC 240

```

Search completed: July 22, 2003, 08:41:17
Job time : 43.1589 secs

```
QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTEESMPNPEGRTYTFGASCTYACP 300
Db 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTEESMPNPEGRTYTFGASCTYACP 300
QY 301 YNYLSTDVGSCTLVCPHLNBOVTAEADGTQCEKCKSPCARVCYGLMFNNFTVSFMR-V 359
Db 301 YNYLSTDVGSCTLVCPHLNBOVTAEADGTQCEKCKSPCARVCYGLMEH-----LREV 353
QY 360 PKYSASHLEEFAGCKKIFGSLAFLPESFGDPPASNTAPLOPEOLQVFETLEETITGYLIS 419
Db 360 PKYSASHLEEFAGCKKIFGSLAFLPESFGDPPASNTAPLOPEOLQVFETLEETITGYLIS 413
QY 420 AMPDLSLPDLSVPQNLQYIRGRILHNGAYSITLQGLGISMGLRSLRELSGLALIHNTH 479
Db 420 AMPDLSLPDLSVPQNLQYIRGRILHNGAYSITLQGLGISMGLRSLRELSGLALIHNTH 473
QY 480 LCFVHTVPMDOLEFNPQALLHTANRPEDECVGEGLAGHQLCARGHGWPSPPTCCVNCSSQ 539
Db 474 LSFVHTVPMDOLEFNPQALLHTANRPEDECVGEGLAGHQLCARGHGWPSPPTCCVNCSSQ 533
QY 540 FLRGQECVEECRVLQGLPREYVNAHCLPCHPECQPNQSVTCFGEADQVACAHYXDP 599
Db 534 FLRGQECVEECRVLQGLPREYVNAHCLPCHPECQPNQSVTCFGEADQVACAHYXDP 593
QY 600 PCVAPRCPSGVKPKPLSYMPIWKFPEDEGACQPCINCTHSCVDLDDKGCFAEQRASPLTS 659
Db 594 PCVAPRCPSGVKPKPLSYMPIWKFPEDEGACQPCINCTHSCVDLDDKGCFAEQRASPLTS 653
QY 660 ILSAVGILLVVLGVVFGILIRROOKIRKTYMRRLQETELVEPLTPSGAMPNQAQMR 719
Db 654 ILSAVGILLVVLGVVFGILIRROOKIRKTYMRRLQETELVEPLTPSGAMPNQAQMR 713
QY 720 ILKETELRKVKVJGSGAFGTYYKGIWIPDGENYKIPYAIKVLRNTSPKANKEILDEAYV 779
Db 714 ILKETELRKVKVJGSGAFGTYYKGIWIPDGENYKIPYAIKVLRNTSPKANKEILDEAYV 773
QY 780 MAGVSPYVSRLLGICLTSTVQCLVTQLMFPGCLLDHRENHRLGSDLLNMCQIAKGM 839
Db 774 MAGVSPYVSRLLGICLTSTVQCLVTQLMFPGCLLDHRENHRLGSDLLNMCQIAKGM 833
QY 840 SYLEDVRLVHDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKXMALES 899
Db 834 SYLEDVRLVHDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKXMALES 893
QY 900 ILRRRFTHOSDVMSYGYVTWELMTFGAKPYDGIAREIPDLLEKGERLPQRPICITIDVYM 959
Db 894 ILRRRFTHOSDVMSYGYVTWELMTFGAKPYDGIAREIPDLLEKGERLPQRPICITIDVYM 953
QY 960 IMVKCMMIDSECRPRFRELVSFERSMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDD 1019
Db 954 IMVKCMMIDSECRPRFRELVSFERSMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDD 1013
QY 1020 MGDLYDAEEYLVQCGFCFDPDPAFGAGVYHHRHSSSTRSGGDIITLGLPSEEDAPRS 1079
Db 1014 MGDLYDAEEYLVQCGFCFDPDPAFGAGVYHHRHSSSTRSGGDIITLGLPSEEDAPRS 1073
QY 1080 PLAPSEGASDVDFDGLGMAKGLQSLPTHDSPLQRYSEDPTVPLPSETDGYVAPLTC 1139
Db 1074 PLAPSEGASDVDFDGLGMAKGLQSLPTHDSPLQRYSEDPTVPLPSETDGYVAPLTC 1133
QY 1140 SPOPEYVNPDPVPPQPSPREGLPARPAGATLERPKTISPGKNGVVDVAFGAVEN 1199
Db 1134 SPOPEYVNPDPVPPQPSPREGLPARPAGATLERPKTISPGKNGVVDVAFGAVEN 1193
QY 1200 PEYLTPGGAAPQPHPPAPSPAFNDLYWDDODPPERGAPOSTFKGTPAENPEYLGIDV 1259
Db 1194 PEYLTPGGAAPQPHPPAPSPAFNDLYWDDODPPERGAPOSTFKGTPAENPEYLGIDV 1253
QY 1260 PV 1261
Db 1254 PV 1255
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 23.0157 Seconds

(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-369-383-14

Perfect score: 6853

Sequence: 1 MELAALCRMGILLALLPPGK.....TRKGTPTANPEYLGIDVPV 1261

Scoring table: BLASTSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: piri: *
2: piri: *
3: piri: *
4: piri: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query	Length	DB ID	Description
1	6655	97.1	1255	1 A24571	protein-tyrosine k
2	5853	85.4	1260	1 TVRNU	protein-tyrosine k
3	5846.5	85.3	1254	1 T48161	p-185 precursor
4	3101	45.3	1210	1 GQHUE	epidermal growth f
5	3075	44.9	1223	1 A53183	epidermal growth f
6	2934.5	42.8	1203	1 TVCHLV	epidermal growth f
7	2934.5	42.8	1308	1 A47253	epidermal growth f
8	2639	38.5	1166	1 S06142	protein-tyrosine k
9	2382.5	34.8	1342	2 A36223	kinase-related tra
10	2284.5	33.5	1339	2 Q4387	epidermal growth f
11	1766.5	25.8	699	1 TVFVLV	protein-tyrosine k
12	1703	24.9	604	1 TVYOH	protein-tyrosine k
13	1647	24.0	544	2 S35745	protein-tyrosine k
14	1640	23.9	545	2 S00727	protein-tyrosine k
15	1623	23.7	540	2 B44776	kinase-related tra
16	1621	23.7	540	1 TVFVEB	protein-tyrosine k
17	1613.5	23.5	1330	1 GQFPE	epidermal growth f
18	1467	21.4	644	2 A36325	epidermal growth f
19	1285	18.8	1333	2 E88257	protein-tyrosine k
20	1285	18.8	1374	2 S70712	protein-tyrosine k
21	1202	17.5	1369	2 S70713	protein-tyrosine k
22	1179	17.2	1377	1 A45558	epidermal growth f
23	1089	15.9	527	2 A42032	epidermal growth f
24	958.5	14.0	843	2 A27331	epidermal growth f
25	806.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.0	311	2 S13808	protein-tyrosine k
27	733	10.7	1363	2 T43200	insulin-like growt
28	698	10.2	1382	1 INHUR	insulin receptor p
29	696	10.2	1607	2 T43212	insulin-like growt

30	691	10.1	1383	2 A36080	insulin receptor p
31	690.5	10.1	1372	2 A34157	insulin receptor p
32	680	9.9	1477	2 T18534	protein-tyrosine k
33	667.5	9.7	1300	2 A36502	insulin receptor-r
34	658	9.6	1268	2 B36502	insulin receptor-r
35	636	9.3	1367	1 IGHURI	insulin-like growt
36	626	9.1	1350	2 T30346	insulin receptor -
37	624	9.1	1371	2 A33837	insulin-like growt
38	619.5	9.0	2148	1 A56081	insulin receptor
39	615	9.0	2101	2 S57245	insulin receptor
40	596.5	8.7	987	2 A54092	protein-tyrosine k
41	588	8.6	1114	1 S05582	protein-tyrosine k
42	586	8.6	1091	2 S33596	protein-tyrosine k
43	585	8.5	976	2 A36355	protein-tyrosine k
44	583.5	8.5	952	2 I50612	protein-tyrosine k
45	581.5	8.5	977	2 S49004	tyrosine kinase Mp

ALIGNMENTS

RESULT 1

A24571 protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N/Alternate names: c-erb-B-2 protein precursor; Kinase-related transforming pr
C/Species: Homo sapiens (man)
C/Date: 25-Oct-1987 #sequence, revision 06-Dec-1996 #text change 11-Jun-1999

C/Date: 25-Oct-1987 #sequence, revision 06-Dec-1996 #text change 11-Jun-1999
R/Accession: A24571; A25491; A44188; B44188; I59509; I57622

R/Accession: T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; S
Nature 319, 230-234, 1986

A/Title: Similarity of protein encoded by the human c-erb-B-2 gene to epiderma
A/Reference number: A24571; M01D:86118653; PMID:3003577

A/Accession: A24571
A/Molecule type: mRNA

A/Residues: 1-1255 <YAM>
A/Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989

R/Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6487-6501, 1985

A/Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB
A/Reference number: A25491; M01D:86016729; PMID:2959567

A/Accession: A25491
A/Molecule type: DNA

A/Residues: 737-1031 <SEM>
A/Cross-references: GB:M11767; NID:g182163; PIDN:AAA35808.1; PID:g553262

R/Consensus, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGarath, J.;
Science 230, 1132-1139, 1985

A/Title: Tyrosine kinase receptor with extensive homology to EGF receptor shar
A/Reference number: A44188; M01D:86070181; PMID:2999974

A/Accession: A44188
A/Molecule type: DNA

A/Residues: 740-910 <COU1>
A/Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989

A/Accession: B44188
A/Molecule type: mRNA

A/Residues: 1-517; 'PALI', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>
A/Cross-references: GB:M11730; NID:g183986

R/King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 228, 974-976, 1985

A/Title: Amplification of a novel v-erbB-related gene in a human mammary carcin
A/Reference number: I59509; M01D:85272597; PMID:2992089

A/Accession: I59509
A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA
A/Residues: 832-909 <REX>

A/Cross-references: GB:I29395; NID:g459807; PIDN:AAA35809.1; PID:g459808
R/Tal, M.; King, C.R.; Kraus, M.H.; Ulrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987

A/Title: Human hHR2 (neu) promoter: evidence for multiple mechanisms for trans
A/Reference number: I57622; M01D:87286898; PMID:3039551

A/Accession: I57622
A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA
A/Residues: 1-191 <PAL>

A:Cross-references: GB:M16792; NID:9183983; PIDN:AA58637.1; PID:9553332
 C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
 C:Genetic:
 A:Gene: GNB:ERBB2; NGL; NEU; HER-2
 A:Cross-references: GDB:120613; OMIM:164870
 A:Map position: 17q21.1-17q21.1
 A:Introns: 25/1; 75/3; 147/1; 883/3
 A>Note: the 1st of introns is incomplete
 C:Function:
 A:Description: catalyzes the phosphorylation of a peptide tyrosine residue by ATP
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase
 F:1-22/Domain: signal sequence #status predicted <SIG>
 F:122-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
 F:122-653/Domain: extracellular #status predicted <EXT>
 F:170-304/Domain: EGF receptor extracellular domain repeat <EB1>
 F:195-605/Domain: EGF receptor extracellular domain repeat <EB2>
 F:554-675/Domain: transmembrane #status predicted <TM>
 F:576-1255/Domain: intracellular #status predicted <INT>
 F:718-983/Domain: protein kinase homology <KIN>
 F:726-734/Region: protein kinase ATP-binding motif
 F:68/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F:753/Active site: Lys #status predicted
 F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 97.1%; Score 6655; DB 1; Length 1255;
 Best Local Similarity 97.4%; Pred. No. 8,662,666;
 Matches 1229; Conservative 11; Mismatches 14; Indels 8; Gaps 2;

```

QY      1 MELALCRWGLLALLPFGASTQVCTGTDMKRLPLSPSPETHLMDLHLYOGCCVVOGNTL 60
DB      1 MELALCRWGLLALLPFGASTQVCTGTDMKRLPLSPSPETHLMDLHLYOGCCVVOGNTL 60

QY      61 ELTYLPINASLSFLDIOEVQGYVLIANQVROYPLQRLTRVGTQLEFENYALAVDNG 120
DB      61 ELTYLPINASLSFLDIOEVQGYVLIANQVROYPLQRLTRVGTQLEFENYALAVDNG 120

QY      121 DPLNNTTPTVGASPGQLRELQRLSTELIKGVLIQRPQICVODTILMKDIFHKNNQLA 180
DB      121 DPLNNTTPTVGASPGQLRELQRLSTELIKGVLIQRPQICVODTILMKDIFHKNNQLA 180

QY      181 LTLIDNRSRACHPSPCKSGRCWGSSESDCQSLTRTVACGAGCARCKGPPFDCCHEQC 240
DB      181 LTLIDNRSRACHPSPCKSGRCWGSSESDCQSLTRTVACGAGCARCKGPPFDCCHEQC 240

QY      241 AAGCTGPKHSDCLAFHNSGICELHCPALVTNTDTFESMPNPEGRTYEGASCVTACP 300
DB      241 AAGCTGPKHSDCLAFHNSGICELHCPALVTNTDTFESMPNPEGRTYEGASCVTACP 300

QY      301 YNYLSTDVGSCTLVQPLINQEVTAEDGQREKSKRCARCYGLGFNNFTVSWIR-V 359
DB      301 YNYLSTDVGSCTLVQPLINQEVTAEDGQREKSKRCARCYGLGFNNFTVSWIR-V 359

QY      360 PYVASHLSEFPAGCKKIFGSLAFLPESFGDPASNTAPQVYKANSKFIQTTELGYIYS 419
DB      360 PYVASHLSEFPAGCKKIFGSLAFLPESFGDPASNTAPQVYKANSKFIQTTELGYIYS 419

QY      420 AMPDSLPLDSVFQNTQVTRGRILHNGAVSLTQGLGISWGLSLRLSLGSLALIHNTH 479
DB      420 AMPDSLPLDSVFQNTQVTRGRILHNGAVSLTQGLGISWGLSLRLSLGSLALIHNTH 479

QY      479 LCFVHTVPMQDLFRPHQALLHTANRPEDEVGSLGCHQICARHGHWGGPQVCVNCQ 539
DB      479 LCFVHTVPMQDLFRPHQALLHTANRPEDEVGSLGCHQICARHGHWGGPQVCVNCQ 539

QY      540 FLRGOECVEECRVYQGLPREYVNAHCLPCHPRECQFQNGSVTCFGEPADQVCAAHYKP 599
DB      540 FLRGOECVEECRVYQGLPREYVNAHCLPCHPRECQFQNGSVTCFGEPADQVCAAHYKP 599

QY      599 PCVCARCPGSKVPLDSYMPIMKPFDEEGACQPCINCTHSCVLDLMDGCAEGRASPPLTS 659
DB      599 PCVCARCPGSKVPLDSYMPIMKPFDEEGACQPCINCTHSCVLDLMDGCAEGRASPPLTS 659

```

DB 594 PCVCARCPGSKVPLDSYMPIMKPFDEEGACQPCINCTHSCVLDLMDGCAEGRASPPLTS 659
 QY 660 IYSAVNGILLVYLVGVVGLIKRQCKIRKRYTKRRLQETELVEPLTPSGAMPNQMR 711
 DB 654 IISAVNGILLVYLVGVVGLIKRQCKIRKRYTKRRLQETELVEPLTPSGAMPNQMR 711
 QY 720 ILKETELRKVKLGSGAGFYKGIWIPDGENVKIPAIKYLRENTSPKANKELIDAVY 771
 DB 714 ILKETELRKVKLGSGAGFYKGIWIPDGENVKIPAIKYLRENTSPKANKELIDAVY 771
 QY 780 MAGVSPYVSRLLGLCTSTQVLTQMLPYGCLLDHYENKRGSLQDILNMCQIAKGM 831
 DB 774 MAGVSPYVSRLLGLCTSTQVLTQMLPYGCLLDHYENKRGSLQDILNMCQIAKGM 831
 QY 840 SYLEVRIVHVDLAARNVLYSPHNVKITDGLARLDDIETETHADGGVPIKMMLES 891
 DB 834 SYLEVRIVHVDLAARNVLYSPHNVKITDGLARLDDIETETHADGGVPIKMMLES 891
 QY 900 ILRRFTQSQDWSYGVTVWELMTFGAKPYDGLPAREIPDLEKGERLPQPICTIDVYM 951
 DB 894 ILRRFTQSQDWSYGVTVWELMTFGAKPYDGLPAREIPDLEKGERLPQPICTIDVYM 951
 QY 960 IMVKCMITDSECRPRFRLVSESSRMARDQRFVIONEDLGPASPLDSTFYSLLEDD 101
 DB 954 IMVKCMITDSECRPRFRLVSESSRMARDQRFVIONEDLGPASPLDSTFYSLLEDD 101
 QY 1020 MODLVDAEYLVPOQGFPCDPAPAGAGMHHRRSSSTSGGDLTLGLSPSEEAPRS 107
 DB 1014 MODLVDAEYLVPOQGFPCDPAPAGAGMHHRRSSSTSGGDLTLGLSPSEEAPRS 107
 QY 1080 PLAPSEGAGSVFPGDLMGAKLQSLPTHPSPLQRYSEDPVPLPSSTDGVVAPLTC 113
 DB 1074 PLAPSEGAGSVFPGDLMGAKLQSLPTHPSPLQRYSEDPVPLPSSTDGVVAPLTC 113
 QY 1140 SPOPEYVQCPVRQPPSPREGRPLPARPGATLTERKTLSPGNGVYKQVAFGAVEN 115
 DB 1134 SPOPEYVQCPVRQPPSPREGRPLPARPGATLTERKTLSPGNGVYKQVAFGAVEN 115
 QY 1200 PEYLPQGAAPQHPPPAPFAPFADNLVYWDODPPEEGAPSTFQPTAENEYGLADV 125
 DB 1194 PEYLPQGAAPQHPPPAPFAPFADNLVYWDODPPEEGAPSTFQPTAENEYGLADV 125
 QY 1260 PV 1261
 DB 1254 PV 1255

RESULT 2
 TYRINT
 protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
 C:Accession: A24562; A61204
 R:Barzmann, C.I.; Hung, M.C.; Weinberg, R.A.
 Nature 319, 226-230, 1986
 A>Title: The neu oncogene encodes an epidermal growth factor receptor-related
 A:Reference number: A24562; MUID:86118662; PMID:3945311
 A:Accession: A24562
 A:Molecule type: mRNA
 A:Residues: 1-1260 <BAR>
 A:Cross-references: EMBL:X03362; NID:956745; PIDN:CAA27059.1; PID:956746
 R:Maui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.
 Carcinogenesis 12, 1975-1978, 1991
 A>Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain re-
 2-thiazolyl]formamide or N-methyl-N-nitrosourea.
 A:Reference number: A61204; MUID:920525293; PMID:1682063
 A:Accession: A61204
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 637-663, 'V', 665-702 <MAS>
 A>Note: authors translated the codon GCA for residue 25 as Val
 C:Genetics:
 A:Gene: neu

C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 F:1-13/Domain: signal sequence #status predicted <SIG>
 F:20-1260/Product: protein-cytosine kinase neu #status predicted <MAT>
 F:658-680/Domain: transmembrane #status predicted <TMN>
 F:723-988/Domain: protein kinase homology <KIN>
 F:731-739/Region: protein kinase ATP-binding motif
 F:721,191,263,555,576,634/Binding site: carboxylate (asn) (covalent) #status predicted
 F:691/Binding site: phosphate (thr) (covalent) #status predicted
 F:758/Active site: Lys #status predicted
 F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 85.4%; Score 5853; DB 1; Length 1260;
 Best Local Similarity 85.7%; Pred. No. 6,5e-233;

Matches 1083; Conservative 58; Mismatches 113; Indels 10; Gaps 4;

```

QY      1 MELAALCRWGLLALLPRGAASVQCTGTDKRLPASPEHLDMLRHLYOGCQVQGNL 60
DB      4 MELAAMCRWGLLALLPRGIAQVCTGTDKRLPASPEHLDMLRHLYOGCQVQGNL 63
QY      61 ELTYLPNASTLSPLODIOEVQVYLIAHNOVRVPIQRIRIVGTOLFEDNYALAVLDNG 120
DB      64 ELTYLPANASTLSPLODIOEVQVYLIAHNOVRVPIQRIRIVGTOLFEDNYALAVLDNR 123
QY      121 DPLNNTTPVT-GASPGGLRELQRLSLTEILKGVLIQRNPQLCYQDTILMKDIFHKNQLA 179
DB      124 DPGQVVAATPGRRPEGLRELQRLSLTEILKGVLIQRNPQLCYQDMVLMKDVFRKNQLA 183
QY      180 ALTLIDTRSRACHPCSPCKGSRCKWBSSEDCOSLTRVYACAGGACRCKPLPTDCCHQ 239
DB      184 APVDIDTRSRACHPCSPCKGSRCKWBSSEDCOSLTRVYACAGGACRCKPLPTDCCHQ 243
QY      240 CAAGCTGPKHSDCLACHFNHSG:CELCPCALVYNTDTESNPNEGRTFGASCVTAC 299
DB      244 CAAGCTGPKHSDCLACHFNHSG:CELCPCALVYNTDTESNPNEGRTFGASCVTAC 303
QY      300 PYNVLTSDVGSCTIVCPILHNEVTAEDGTORCEKSCPCARVQVGLGMENFTVSEFMLEV 359
DB      304 PYNVLTSDVGSCTIVCPILHNEVTAEDGTORCEKSCPCARVQVGLGMENFTVSEFMLEV 356
QY      360 PK-VASHLKEEFACKCKIFGSLAEPSPGDDPASNTAPQVYIANKSKFQITLTGLYLI 418
DB      357 ARAITSDVQEFDDCKKIFGSLAEPSPGDDPASNTAPQVYIANKSKFQITLTGLYLI 416
QY      419 SAWPDSLPLDSVFQNLIRIRILHNGAVSLTQGLISWLGRLSRLREGSGALIHNT 478
DB      417 SAWPDSLPLDSVFQNLIRIRILHNGAVSLTQGLISWLGRLSRLREGSGALIHNT 476
QY      479 HLCVHTVPMQDLFRNPHQALLTANRPEDC-CVBSGLACHQLCARGHCWGSPGTQVCNC 537
DB      477 HLCVHTVPMQDLFRNPHQALLHSGNRPEBDLCVSSGLVCNLSLCAHGHGCGPPTQVCNC 536
QY      538 SQFLRGOECVECEGVLLQGLPREYVNAHRCPCHECQRPONGSVTCGRPADQCVAAHAK 597
DB      537 SHFLRGOECVECEGVWGLPREYVSDKRCPCHECQRPONGSVTCGRPADQCVAAHAK 596
QY      598 DPFPCVACRPSGVKPDISYMPWKFPDEGACQPCPCINCTHSCVDDLDKCCAPAEQASPL 657
DB      597 DSSCVACRPSGVKPDISYMPWKFPDEGACQPCPCINCTHSCVDDLDKCCAPAEQASPL 656
QY      658 TSISAVAVGILLVVLGVNFGILLIKRQOKIRKYTRRLLQSTELVEPLTPSAMNQAQ 717
DB      657 TSISAVAVGILLVVLGVNFGILLIKRQOKIRKYTRRLLQSTELVEPLTPSAMNQAQ 716
QY      718 MRILKETELRKVYLGSAGFVYKGIWIPDGENVKI PVAIKYLRENTSPKANKELIDEA 777
DB      717 MRILKETELRKVYLGSAGFVYKGIWIPDGENVKI PVAIKYLRENTSPKANKELIDEA 776
QY      778 YVMAGVSPYVSRLLGLCTSTVQVLTQMLPYGCLLDHRENNRGSGODLLNMCQIAK 837
DB      777 YVMAGVSPYVSRLLGLCTSTVQVLTQMLPYGCLLDHRENNRGSGODLLNMCQIAK 836
QY      838 GMSYLEDVRLVHRDLAARNVLYKSPNHVKITDFTGLARLLDIDETEVHADGKVPKIMMAL 897

```

```

DB      837 GMSYLEDVRLVHRDLAARNVLYKSPNHVKITDFTGLARLLDIDETEVHADGKVPKIMMAL 896
QY      898 ESILRRRTTHSDVMSYGVYTWELMTFGANPYDGI PARSEIPDLLEKERLPPOPICTIDV 957
DB      897 ESILRRRTTHSDVMSYGVYTWELMTFGANPYDGI PARSEIPDLLEKERLPPOPICTIDV 956
QY      958 YMIWKCKMIDSECRPPRELVEFSFMAWDPORFVYIQNEDUGPASPLDSTYRSLLED 101
DB      957 YMIWKCKMIDSECRPPRELVEFSFMAWDPORFVYIQNEDUGPASPLDSTYRSLLED 101
QY      1018 DDMGDLVAEEYLPQOGFPCDDAPAGGMVHRRHSSSTRSGGDLTLGLEPSEEAR 107
DB      1017 DDMGDLVAEEYLPQOGFSPDPTPGTSTARRHSSSTRSGGDLTLGLEPSEEAR 107
QY      1078 RSLPASEGASDVFDGDLGKAAGLQSLPTHDPSPLQRYSEDPVPLPSETDGYAPL 113
DB      1077 RSLPASEGASDVFDGDLGKAAGLQSLPTHDPSPLQRYSEDPVPLPSETDGYAPL 113
QY      1138 TCSQRPPEYVQPDVRFQPSRREGPLPAAPPAQTLEKATLSFGKXGVYKDVFAFGAV 115
DB      1137 ACSQRPPEYVQPDVRFQPSRREGPLPAAPPAQTLEKATLSFGKXGVYKDVFAFGAV 115
QY      1198 ENPEYLPFOGGAAPQHPFPAFSPFNLYYMOQDPPERGAPSTFGPTAENPEYLG 125
DB      1197 ENPEYLPFBGTAAPFPAFSPFNLYYMOQDPPERGAPSTFGPTAENPEYLG 125
QY      1258 DVPV 1261
DB      1257 DVPV 1260

RESULT 3
148161
P-185 precursor - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C:Accession: 148161
R:Nakamura, T.; Uehijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.;
Gene 140, 261-255, 1994
A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A:Reference number: 148161; MUID:94193007; PMID:7908275
A:Accession: 148161
A:Status: preliminary; translated from GB/EMBL/DDJ
A:Molecule type: mRNA
A:Residues: 1-1254 <RES>
A:Cross-references: GB:D16295; NID:9493236; PIDN:BA03801.1; PID:9747595
C:Genetics:
A:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif

Query Match 85.3%; Score 5846.5; DB 2; Length 1254;
Best Local Similarity 85.3%; Pred. No. 1.2e-232;
Matches 1077; Conservative 69; Mismatches 107; Indels 9; Gaps 3

```

```

QY      1 MELAALCRWGLLALLPRGAASVQCTGTDKRLPASPEHLDMLRHLYOGCQVQGNL 60
DB      1 MELAAMCRWGLLALLPRGIAQVCTGTDKRLPASPEHLDMLRHLYOGCQVQGNL 60
QY      61 ELTYLPNASTLSPLODIOEVQVYLIAHNOVRVPIQRIRIVGTOLFEDNYALAVLDNG 120
DB      61 ELTYLPANASTLSPLODIOEVQVYLIAHNOVRVPIQRIRIVGTOLFEDNYALAVLDNR 120
QY      121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGVLIQRNPQLCYQDTILMKDIFHKNQLA 180
DB      121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGVLIQRNPQLCYQDTILMKDIFHKNQLA 180
QY      181 LTLIDTRSRACHPCSPCKGSRCKWBSSEDCOSLTRVYACAGGACRCKPLPTDCCHQ 240
DB      181 PVDIDTRSRACHPCSPCKGSRCKWBSSEDCOSLTRVYACAGGACRCKPLPTDCCHQ 240

```


QY 241 AAGCTGPKASDCLACLHFNHSGICELHCPALVTYNTDFFESMPNDEGRYTGASCYACP 300
 DB 241 AAGCTGPKASDCLACLHFNHSGICELHCPALVTYNTDFFESMPNDEGRYTGASCYACP 300
 QY 301 VNYLSTDVSSCTLVCPPLNHNQEVTAADGQCEKCKSPCARVCYGLGPNFTYSFWLRVP 360
 DB 301 VNYLSTDVSSCTLVCPPLNHNQEVTAADGQCEKCKSPCARVCYGLGPNFTYSFWLRVP 360
 DB 301 VNYLSTDVSSCTLVCPPLNHNQEVTAADGQCEKCKSPCARVCYGLGPNFTYSFWLRVP 360
 QY 361 K-VASASHEEPAGCKKIFGSLAFLEPESFDGPASNTAPQYIKANSKFIQITELTGYIYS 419
 DB 361 K-VASASHEEPAGCKKIFGSLAFLEPESFDGPASNTAPQYIKANSKFIQITELTGYIYS 419
 QY 420 AMPDSLPLDSVFQNTQVIRGRILNNGAVSLTLQGLISWLGRLSRLRELGSLALIHNTH 479
 DB 420 AMPDSLPLDSVFQNTQVIRGRILNNGAVSLTLQGLISWLGRLSRLRELGSLALIHNTH 479
 QY 480 LCFVHTVPMWDLFRNPHQALLHTARPEDECVBESLACHQICAGHCHWGPPTQCVNCSQ 539
 DB 480 LCFVHTVPMWDLFRNPHQALLHTARPEDECVBESLACHQICAGHCHWGPPTQCVNCSQ 539
 QY 474 LCFVHTVPMWDLFRNPHQALLHTARPEDECVBESLACHQICAGHCHWGPPTQCVNCSH 533
 DB 474 LCFVHTVPMWDLFRNPHQALLHTARPEDECVBESLACHQICAGHCHWGPPTQCVNCSH 533
 QY 540 FLRGCEVCECRVLQGLPREVYNARHCLPCHPECOPOGNGSYTCFGEPADQVACAHYKDP 599
 DB 540 FLRGCEVCECRVLQGLPREVYNARHCLPCHPECOPOGNGSYTCFGEPADQVACAHYKDP 599
 QY 534 FLRGCEVCECRVLQGLPREVYNARHCLPCHPECOPOGNGSYTCFGEPADQVACAHYKDS 593
 DB 534 FLRGCEVCECRVLQGLPREVYNARHCLPCHPECOPOGNGSYTCFGEPADQVACAHYKDS 593
 QY 600 PECVACRCSGVKPDLSYMPIMKFPBESACQPCPINCCHSCVDIDDKGCPABOASPLTS 659
 DB 600 PECVACRCSGVKPDLSYMPIMKFPBESACQPCPINCCHSCVDIDDKGCPABOASPLTS 659
 QY 594 PECVACRCSGVKPDLSYMPIMKFPBESACQPCPINCCHSCVDIDDKGCPABOASPLTS 653
 DB 594 PECVACRCSGVKPDLSYMPIMKFPBESACQPCPINCCHSCVDIDDKGCPABOASPLTS 653
 QY 660 IYSAVVGILLVAVLVEFGILIKRROQKIRKYTWMLQETELVEPIPPSGAMNQAKMR 719
 DB 660 IYSAVVGILLVAVLVEFGILIKRROQKIRKYTWMLQETELVEPIPPSGAMNQAKMR 719
 QY 654 IYATVVGILLVAVLVEFGILIKRROQKIRKYTWMLQETELVEPIPPSGAMNQAKMR 713
 DB 654 IYATVVGILLVAVLVEFGILIKRROQKIRKYTWMLQETELVEPIPPSGAMNQAKMR 713
 QY 720 IYKTELKRVKVLGSGAGFTYKGIWIDGSENVKI PVAIKYLRENTS PKANKELIDRAYV 779
 DB 720 IYKTELKRVKVLGSGAGFTYKGIWIDGSENVKI PVAIKYLRENTS PKANKELIDRAYV 779
 QY 714 IYKTELKRVKVLGSGAGFTYKGIWIDGSENVKI PVAIKYLRENTS PKANKELIDRAYV 773
 DB 714 IYKTELKRVKVLGSGAGFTYKGIWIDGSENVKI PVAIKYLRENTS PKANKELIDRAYV 773
 QY 780 MAGVSPYVSRLLGICLTSTVQVLYQMLPACCLDHRNENGRIGSODLLMCKQIAKGM 839
 DB 780 MAGVSPYVSRLLGICLTSTVQVLYQMLPACCLDHRNENGRIGSODLLMCKQIAKGM 839
 QY 774 MAGVSPYVSRLLGICLTSTVQVLYQMLPACCLDHRNENGRIGSODLLMCKQIAKGM 833
 DB 774 MAGVSPYVSRLLGICLTSTVQVLYQMLPACCLDHRNENGRIGSODLLMCKQIAKGM 833
 QY 840 SYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEVHADGKVPKIMALES 899
 DB 840 SYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEVHADGKVPKIMALES 899
 QY 834 SYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEVHADGKVPKIMALES 893
 DB 834 SYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEVHADGKVPKIMALES 893
 QY 900 ILRRPFTHOSDVWYSGVYVWELMTGAPYGCIPAREIPDLLEKGERLPQPPICITIDVYM 959
 DB 900 ILRRPFTHOSDVWYSGVYVWELMTGAPYGCIPAREIPDLLEKGERLPQPPICITIDVYM 959
 QY 894 ILRRPFTHOSDVWYSGVYVWELMTGAPYGCIPAREIPDLLEKGERLPQPPICITIDVYM 953
 DB 894 ILRRPFTHOSDVWYSGVYVWELMTGAPYGCIPAREIPDLLEKGERLPQPPICITIDVYM 953
 QY 960 IMVCKMMIDSECRPFRELVESEFSMARDPQRFVVI QNEDLGPPASPLDSTFYRSLLEDD 1019
 DB 960 IMVCKMMIDSECRPFRELVESEFSMARDPQRFVVI QNEDLGPPASPLDSTFYRSLLEDD 1019
 QY 954 IMVCKMMIDSECRPFRELVESEFSMARDPQRFVVI QNEDLGPPASPLDSTFYRSLLEDD 1013
 DB 954 IMVCKMMIDSECRPFRELVESEFSMARDPQRFVVI QNEDLGPPASPLDSTFYRSLLEDD 1013
 QY 1020 MGDVDADEBYLVPOGFCPCPDPAAGAMVTHRRSSSTRSGGDLTLG:LEPSEEARPS 1079
 DB 1020 MGDVDADEBYLVPOGFCPCPDPAAGAMVTHRRSSSTRSGGDLTLG:LEPSEEARPS 1079
 QY 1014 MGDVDADEBYLVPOGFCPCPDPAAGAMVTHRRSSSTRSGGDLTLG:LEPSEEARPS 1073
 DB 1014 MGDVDADEBYLVPOGFCPCPDPAAGAMVTHRRSSSTRSGGDLTLG:LEPSEEARPS 1073
 QY 1080 PLAPSEGAAGVDFDGLGMAAGKGLSPTHDPSPLORYSEDPVLPSPSTEDVYVAPLTC 1139
 DB 1080 PLAPSEGAAGVDFDGLGMAAGKGLSPTHDPSPLORYSEDPVLPSPSTEDVYVAPLTC 1139
 QY 1074 PLAPSEGAAGVDFDGLGMAAGKGLSPTHDPSPLORYSEDPVLPSPSTEDVYVAPLTC 1133
 DB 1074 PLAPSEGAAGVDFDGLGMAAGKGLSPTHDPSPLORYSEDPVLPSPSTEDVYVAPLTC 1133
 QY 1140 SPQPEYVQPDVRRPQSPSPREGRPLPAAPAGATLERAKTSLPGKNGVYKQVFAFGAVEN 1199
 DB 1140 SPQPEYVQPDVRRPQSPSPREGRPLPAAPAGATLERAKTSLPGKNGVYKQVFAFGAVEN 1199
 QY 1134 SPQPEYVQPDVRRPQSPSPREGRPLPAAPAGATLERAKTSLPGKNGVYKQVFAFGAVEN 1193
 DB 1134 SPQPEYVQPDVRRPQSPSPREGRPLPAAPAGATLERAKTSLPGKNGVYKQVFAFGAVEN 1193
 QY 1200 PEYLTPOGGAAPQPPPPAPFAPADNLVYWDODPEEGAPSTKQPTLENREYVGLDV 1259
 DB 1200 PEYLTPOGGAAPQPPPPAPFAPADNLVYWDODPEEGAPSTKQPTLENREYVGLDV 1259
 QY 1194 PEYLTPOGGAAPQPPPPAPFAPADNLVYWDODPEEGAPSTKQPTLENREYVGLDV 1252
 DB 1194 PEYLTPOGGAAPQPPPPAPFAPADNLVYWDODPEEGAPSTKQPTLENREYVGLDV 1252
 QY 1260 PV 1261
 DB 1253 PV 1254

RESULT 4
 GENE
 epidermal growth factor receptor precursor - human
 N/contamin: protein-tyrosine kinase (EC 2.7.1.112) erbB
 C/species: Homo sapiens (man)
 C/date: 15-Nov-1984 #sequence revision 27-Nov-1985 #text change 11-Jun-1999
 C/accession: A00641; A25772; S30024; A38672; A00642; A43615; A23062; A05281; /
 R/Jillich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.;
 rg, P.H.
 Nature 309, 418-425, 1984
 A>Title: Human epidermal growth factor receptor cDNA sequence and aberrant ex
 A/Reference number: A00641; MUID:84219729; PMID:6328312
 A/Accession: A00641
 A/Molecule type: mRNA
 A/Residues: 1-1210 <URL>
 A/Cross-references: EMBL:X00589; NID:g31113; PIDN:CAA5240.1; PID:g757924
 A/Note: the authors translated the codon AAG for residue 540 as Asn
 R/ishi, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
 Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
 A>Title: Characterization and sequence of the promoter region of the human ep
 A/Reference number: A25772; MUID:85270438; PMID:2991899
 A/Accession: A25772
 A/Status: translation not shown
 A/Molecule type: DNA
 A/Residues: 1-29 <ISH>
 A/Cross-references: GB:M11234; NID:g181981; PIDN:AA52370.1; PID:g553272
 R/Haley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ullrich, A.; Waterfield
 Oncogene Res. 1, 375-396, 1987
 A>Title: The human EGF receptor gene: structure of the 110 kb locus and ident
 A/Reference number: S30024; MUID:86217333; PMID:3329716
 A/Accession: S30024
 A/Molecule type: DNA
 A/Residues: 1-29 <HA2>
 A/Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
 R/Haley, J.D.; Waterfield, M.D.
 J. Biol. Chem. 266, 1746-1753, 1991
 A>Title: Contributory effects of de Novo transcription and premature transcrip
 A/Reference number: A38672; MUID:91107677; PMID:1998448
 A/Accession: A38672
 A/Molecule type: DNA
 A/Residues: 1-29 <HA2>
 A/Cross-references: GB:M38425; NID:g181977; PIDN:AA63171.1; PID:g553271
 A/Experimental source: carcinoma cell line A431.7
 R/Xu, Y.; Iishi, S.; Clark, A.D.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe,
 Nature 309, 806-810, 1984
 A>Title: Human epidermal growth factor receptor cDNA is homologous to a variet
 A/Reference number: A00642; MUID:84245835; PMID:6330563
 A/Accession: A00642
 A/Molecule type: mRNA
 A/Residues: 1-29 <HA2>
 A/Cross-references: GB:M38425; NID:g181977; PIDN:AA63171.1; PID:g553271
 A/Experimental source: carcinoma cell line A431.7
 R/Xu, Y.; Iishi, S.; Clark, A.D.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe,
 Nature 309, 806-810, 1984
 A>Title: Human epidermal growth factor receptor cDNA is homologous to a variet
 A/Reference number: A00642; MUID:84245835; PMID:6330563
 A/Accession: A00642
 A/Molecule type: mRNA
 A/Residues: 150-187, 'KSVIQNV', 195, 'W', 197, 'A', 199-222, 'S', 224-304, 'R',
 '798-799, 'TD', 802-811, 'R', 813-942 <XUY>
 A/Experimental source: A431 human carcinoma cells, which have large numbers of
 R/Lin, C.R.; Chen, W.S.; Krutiger, W.; Stolarek, L.S.; Weber, W.; Evans, R.M.;
 Science 224, 843-848, 1984
 A>Title: Expression cloning of human EGF receptor complementary DNA: gene ampl
 A/Reference number: A43615; MUID:84196372; PMID:6326261
 A/Accession: A43615
 A/Molecule type: mRNA
 A/Residues: 713-964 <LIN>
 A/Experimental source: epidermoid carcinoma cell line A431
 R/Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malle
 Biochem. Biophys. Res. Commun. 124, 125-132, 1984
 A/Reference number: A23062; MUID:85046483; PMID:6093780
 A/Accession: A23062
 A/Molecule type: mRNA
 A/Residues: 1028-1210 <SIM>
 R/Weber, W.; Gill, G.N.; Speiss, J.
 Science 224, 294-297, 1984
 A/Reference number: A05281; MUID:84172183; PMID:6324343
 A/Accession: A05281
 A/Molecule type: Protein
 A/Residues: 25-30, 'S', 32-51, 454-467 <WEB>

R/Russo, M.W.; Lukas, T.J.; Cohen, S.; Stancos, J.V.
 J. Biol. Chem. 260, 5205-5208, 1985
 A/Title: Identification of residues in the nucleotide binding site of the epidermal growth factor receptor
 A/Reference number: A60143; MUID:85182650; PMID:2985580
 A/Accession: A60143
 A/Molecule type: Protein
 A/Residues: 740-744, 'X', 746-747 <RUS>
 R/Mrozowski, B.; Nosiq, G.; Cohen, S.
 Nature 309, 270-273, 1984
 A/Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide dismutase
 A/Reference number: A38023; MUID:84191554; PMID:6325948
 A/Accession: A38023
 A/Contents: annotation: receptor activity
 A/Note: The EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
 R/Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsch, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
 Cell 59, 33-43, 1989
 A/Title: Functional independence of the epidermal growth factor receptor from a domain 1
 A/Reference number: A3331; MUID:90003233; PMID:2790960
 A/Accession: A3331
 A/Contents: annotation: internalization signal
 C/Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor complex
 A/Genes: GDB:EGFR
 A/Cross-references: GDB:120610; OMIM:131550
 A/Map position: 7p12.3-7p12.1
 C/Superfamily: epidermal growth factor receptor; protein kinase homology
 C/Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphotyrosine; signal sequence #status predicted <SIG>
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-1210/Product: EGF receptor #status predicted <MAT>
 F:25-645/Domain: extracellular #status predicted <EXT>
 F:75-300/Domain: EGF receptor extracellular domain repeat <EE1>
 F:390-600/Domain: EGF receptor extracellular domain repeat <EE2>
 F:668-668/Domain: transmembrane #status predicted <TM>
 F:669-1210/Domain: intracellular #status predicted <INT>
 F:710-975/Domain: protein kinase homology <KIN>
 F:718-726/Region: protein kinase ATP-binding motif
 F:999-1046/Region: coated-pit mediated internalization signal
 F:1247-1250/Region: inhibitory
 F:128,175,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:745/Active site: Lys #status experimental

Query Match 45.3%; Score 3101; DB 1; Length 1210;
 Best Local Similarity 49.3%; Pred. No. 4e-120;
 Matches 626; Conservative 175; Mismatches 358; Indels 112; Gaps 23;

```

QY 11 LLLALLPPGAA--STVCTGTDMKLLPASPETHLDMLEHLYOGGVQVQNELTYLPTN 68
DB 14 LLAALCPASALEEKVKVCGSTSKLQLGTPEHDHLSLORMENNEVGLNLEIITYARN 73
QY 69 ASLSFLQDIOEVGVYLIANOVROYPLQRLIVGTQLFEDNYLAVLDNGDPIINNTTP 128
DB 74 YDLSFLKTIQEVAVGVYLIANTVERPLENLIQIRGMYYENSVALAVLSNYD----- 126
QY 129 VTGASPGGLRELDLSLTLEIKGVLIQENPOLCYODITLWMDIFHKNNQLATLIDTNR 188
DB 127 ---ANKTGKLELPMKRLQELHCAVAFSNPNALCNVESIQMDIVSSPLSNMSPDNH 183
QY 189 SRACHPCSPWCKSRGWGSSSDQSLRTVCAGCA-RCKGPLRTDCHDECAAGCTGP 247
DB 184 LGSCKCKDPSCPGSCWAGGEECCQLTKIICAOQCSGCRGKSPDSCCHQCAAGCTGP 243
QY 248 KHSDCLACLHFNHSGICELHCPALVTYNTDFESMPNPEGRTRFASCVTACPNYISTD 307
DB 244 RESDCLVCKRFDEATCKDTCPLMLYNTFTYQMDVNPGRKSFATVKKCFRRYVYTD 303
QY 308 VGSCTLVCPHLNOEVTAEADGTQRCCKSPCARVCYGLGMENFTVSWLVPKVASHL 367
DB 304 HGSCTVACGADSYEM-EEDGVYKCKCKCEPCKKVCNGIGI-GEFDSL-----SINATNI 356
QY 368 EEFAGCKTIFGSLAFIPESFPDDPASNTAPQIKANSKFTIGTLETGLYISAMPDSIPD 427
DB 357 KHFKNCTSIISGDHILIPVAFRQDSFTHPPPLDPOELDLTKVKEITGLLQAMPENKTD 416
QY 428 LSVFQNLQVIRGLIHNGAVSLTQGLIGISLGLSRELGSGLLIHHNTHLCFVHTVP 487

```

```

DB 417 LHAENLEIRGTRKHQGFSLAVSNLNTSLGRSLKEISDGVYISGNKLCYANTIN 476
QY 488 MDQLFRPHQALHTANRPDECEVSGGLACHQCAHGMCGGPGPTQCVNCSQFLRQOECV 547
DB 477 WKTLFGISGKTKIISNKGENSEKATGVCHALCSFSGCWGEBRDCVSCRVNRSRECV 536
QY 548 EECRVYQGLFREYVNAHCLPCHPECOPONGSVTCFGEPADOCVACAHYKDPFVCVARP 607
DB 537 DKCKLLEGREFREVENSECTIQCHPECLPQAMNITCTGPRDNCIQCAHYIDGPHCVKTCR 596
QY 608 SGYKPLDSVYMPKFPDEBACQPCPGINTHSCVLDLDDGCAEGRASLSTIVSNVNG 666
DB 597 AGVWGENNTL-VKTVADAGHVCILCPNCTCYCTGCTGCGLECPNPKIP--SIAGWGA 653
QY 667 --LLVVLGVVFGILIKRQOKIRKYTRRLLOETELVEPLTPSGAMPNQAKRILKET 724
DB 654 LLLLVVVALGIG---LFMRRLHVRKTRRLLORELVEPLTPSGEARNQALLILKET 710
QY 725 ELARKVYLSGAGFYTKGIWIPDGENVKIPVATKVLRENTSPKANKELDEAVYMAVG 784
DB 711 EFKKIKVLGSGAFGVYKGLMIPGEKVKIPVAIKELREATSPKANKELDEAVYMASVD 770
QY 785 SPYVSRLLGICLSTVQVLTQAMPYGCILDHYRENRGRIGSQDLINWCMQIAKMSYIED 844
DB 771 NPHVCRLLGICLSTVQVLTQAMPYGCILDHYRENRGRIGSQDLINWCMQIAKMSYIED 830
QY 845 VRLVHDLAARNVLYKSPHVKITDFGLARLLDIDETEXHADGKGVPIKMALESILRR 904
DB 831 RRLVHDLAARNVLYKSPHVKITDFGLARLLDIDETEXHADGKGVPIKMALESILRR 890
QY 905 FTHQSVWAGVYVWELMTFGAKRPYDGIIPAREIIPDLLEKGERLPPPICTIDVYIMVKC 964
DB 891 YTHQSVWAGVYVWELMTFGAKRPYDGIIPAREIIPDLLEKGERLPPPICTIDVYIMVKC 950
QY 965 WMTDSECRPFRELESEFARMARDPQRFVYIQ-NEDLGPASPLDSTFRSLDEDDMGDL 102
DB 951 WMTDASRKRFRELESEFARMARDPQRFVYIVIGDGRMLPSTDSNFRALMDEDDMDV 1014
QY 1024 VDAEYLVYPOQGFCCDPAAPAGAGVNHRRSSSTRSGGGLTGLLEPSEEBAPSPAP 108
DB 1011 VDAEYLVYPOQGFCCDPAAPAGAGVNHRRSSSTRSGGGLTGLLEPSEEBAPSPAP 1034
QY 1084 SEGAGSDVDFDGLGMAAKGLQSLPHDPSLPQRYSEDPVYLPSET--DGVAALTCSP 114
DB 1037 SLSATSN--NSTVACIDRNGLOSCEPKEDSFLQRYSSDPTGALTDSIDDTFL-----P 1086
QY 1142 QPEYVNPDPVRQPPSPREGRPLPAARPAAGATLERAKTLPGRKGVYKDVFAFGAVENPE 1201
DB 1089 VPEYINQ-SVPRKPAQSVQNPVYHNPINP-----APSRDHYQD--PHSTAVGNPE 1137
QY 1202 YL-TPQGGAAQPHPPAPSPAFNDLYWQ-----DB-----PERGAPSTFK 1244
DB 1138 YLNTVQ-----PTCVNSTFDSPPAHMAQKSHQISLDNPDYQDFFPEAKENGIFK 1186
QY 1245 GTFEAPNPEYL 1255
DB 1189 GS-TAENAYYL 1198

```

RESULT 5
 A53183
 Epidermal growth factor receptor precursor - mouse
 C/Species: Mus musculus (house mouse)
 C/Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999
 C/Accession: A53183; A43818; S24942; A28941; S45325; I49643
 R/Luetteke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkin
 Genes Dev. 8, 399-413, 1994
 A/Title: The mouse waved-2 phenotype results from a point mutation in the EGF r
 A/Reference number: A53183; MUID:94170986; PMID:8125255
 A/Accession: A53183
 A/Molecule type: mRNA
 A/Residues: 1-1210 <LUE>
 A/Cross-references: GB:U03425

R:Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.
 Oncogene 6, 673-676, 1991
 A:Title: Comparison of EGF receptor sequences as a guide to study the ligand binding site
 A:Reference number: A43818; MUID:91232866; PMID:2050916
 A:Accession: A43818
 A:Molecule type: mRNA
 A:Residues: 1-714 <AVI>
 A:Cross-references: GB:X59698
 R:Elisinger, D.P.; Serrero, G.
 submitted to the EMBL Data Library, June 1992
 A:Reference number: S24942
 A:Accession: S24942
 A:Molecule type: mRNA
 A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>
 A:Cross-references: EMBL:Z12608
 R:Heisermann, G.J.; Gill, G.N.
 J. Biol. Chem. 263, 13152-13158, 1988
 A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
 A:Reference number: A28941; MUID:80330814; PMID:3138233
 A:Accession: A28941
 A:Molecule type: protein
 A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009,
 R:Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
 submitted to the EMBL Data Library, April 1994
 A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
 A:Reference number: S45325
 A:Accession: S45325
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-971, 'K', 973-1210 <VER>
 A:Cross-references: EMBL:X78987; NID:G488830; PIDN:CAA55587.1; PID:G488831
 R:Patla, B.C.; Das, S.K.; Andrews, G.K.; Day, S.K.
 Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
 A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse B
 A:Reference number: I49643; MUID:93126380; PMID:7678348
 A:Accession: I49643
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 12-20, 22-132 <RES>
 A:Cross-references: GB:L06864; NID:G193001; PIDN:AAA51029.1; PID:G567201
 C:Genetics:
 A:Gene: EGFR
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphoproc
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:648-677/Domain: transmembrane #status predicted <TM>
 F:712-977/Domain: protein kinase homology <KIN>
 F:720-728/Region: protein kinase ATP-binding motif
 F:680, 695/Binding site: phosphate (Thr) (covalent) #status experimental
 F:697, 1070, 1071/Binding site: phosphate (Ser) (covalent) #status experimental
 F:693/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
 F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
 F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 44.98; Score 3075; DB 2; Length 1210;
 Best Local Similarity 49.28; Pred. No. 4, 7e-119;
 Matches 630; Conservative 169; Mismatches 361; Indels 120; Gaps 26;

Qy	11	LTALLPFGAA--STOYCTGTDMKLRLPASPEHLDMLRHLYGCGQVQGNILETTLPLN	68
Db	14	LTTLCAAGGALBEKKKCGGTNSRLTOLGTFEDHFLSLQRMNNCCVAVGNILETTLVQGN	73
Qy	69	ASISFLQDIOEVQGYVLIANQVRLQRLIVRGTOLEFEDNYVALAVDNGDPNNTTP	128
Db	74	YDLSFLKTIQEVAGYVLIANTVERPLENLQIIRGNALYENTYALATLSN-----	124
Qy	129	VTGASPGGLRELQRLSTELLKGGVLIQRNPQCYODTILMKDI---FKKNQLALTLI	184
Db	125	YGTNRGLRELPKRNQETLIGAVFSSNPILCNMDITVQNVKSNKSNWDL---	180
Qy	185	DYNSRACHPCSPKCKSCRCMGESSEPCQSLRTVCGAGCA-RCKGPLPTDCCHQCAAG	243
Db	181	OSHPSCPKCDPECPGSCMGGEENCGQLTKIICAGQCSHRCRGRSPSSDCNHCAGG	239

Qy	244	CTGPKSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRTFGASCYACPVNY	303
Db	240	CTGRSEDCLVCCQFQDEATCKTCPLMLYNTTYQMVNPGKYSFGATCYKCKPRNY	299
Qy	304	LSTDVGSCTLVCPLNQNEVTAEDGTORCEKSKPCARVCYGLGM--FNNFTVSWLRVPK	361
Db	300	VYTDHGSQVACGPDYEV-EEGDIKCKCKCDGPRCKVCNGDIGFCKD-TLS-----	350
Qy	362	VSASHLEEPGCKKIFGSLAFLEPESDGPASTAQYKANSKFIGITELGYLISGM	421
Db	351	INATNIGFFKCYCAISGDLILFVAKFGSFTTPPLDPRELEILKTVEITGFLILQAM	410
Qy	422	PDSLPDSVFQNLQVIRGLIHNGAVSLTLOGIGIMLGRSLREKSGSLIHNHTHC	481
Db	411	PDNNTDLHAFENLEIRGTRKQNGSLAVGNTSLGRSLKESISDGVIISSGRNIC	470
Qy	482	FVHTVPMQDLPNPHQALLHTANRPEDECVGEGLACHQLCARGHCKGPPPTCCVNSQFL	541
Db	471	YANTINMKKLFQTPNQTKIMNNAEKDCAVHNVGNPLCSSGCGPFPDCCVSCONVS	530
Qy	542	RGQECVECEVLQGLPREYVNAHCLPCHPEOQPOKGSVTCGPRADOCVACAHYKDPF	601
Db	531	RGKCEVCKNILEBPREFVENSECYQCHPECLPQAMNITCTGRGPDNCTICAHYIDGFH	590
Qy	602	CVARCPGVKPDLSYMPIMKFPDEGACQPCPINCTHSCVDLDDKXCPAEQASPLTSIV	661
Db	591	CVKTCFPAIGENNTL-VMKYADANNVCHLCHANCTYGACAGELQCEVWBSGPKIPSLA	649
Qy	662	SAVIGILLVVLGVVGI-LIKRQOKIKRYTMRLLQETELVEPLTBSGAMPNOQMKI	720
Db	650	TGVLGGLLFTIV-VAGIGLPMRRRIHVKTILRLLOERELEVEPLTBSGEAPNOQHLRI	708
Qy	721	LKETELKRVKVLGSGAFGYVYKIMIPDENVYIPVAKVLENTSPKANKIILDAVYM	780
Db	709	LKEREFKKIVLDSGARFGYTKLMIPBEKXKIPVAILKELNATSPKANKIILDAVYM	768
Qy	781	AGVSPVYSRLGICLTSTVQVLTQMLPFGCLLDVHNRGRGLSGQDLLNMCQIAKGS	840
Db	769	ASVDNPNVCHRLGICLTSTVQVLTQMLPFGCLLDVHNRGRGLSGQDLLNMCQIAKGN	828
Qy	841	YLEPVRVNRDLAARVLYKSPNHYKITPFGALRIDETEYHNDGGVVPKMMALESI	900
Db	829	YLEPVRVNRDLAARVLYKTPPHVKITPFGALKIGABEKYHAGGVVPMKMALESI	888
Qy	901	LRRRFTQSDVMSYGVYVLELMTFGAKPYDGIIPAREIPDLEKGERLPQPPCTIDVYMI	960
Db	889	LHRYTHQSDVMSYGVYVLELMTFGSKPYDGIIPASDISILEKGERLPQPPCTIDVYMI	948
Qy	961	MYCKMTIDSCRRREFELVSEFMRMRDQRFVIO-NDLGPASPLDSTFRSLLEDD	101
Db	949	MYCKMTIDASRREFELVSEFMRMRDQRFVIO-NDLGPASPLDSTFRSLLEDD	100
Qy	1020	MGDLVDAEYLVPOQGFPCDPAPGAGGVHNRHRSSTRSGGDLTLLEPSEEPAPS	107
Db	1009	MEDVDADEVLTQQGF-----NSPST-----SRT	103
Qy	1080	PLAPSEAGSDVDGDLGMAKGLQSLFTHDPSPLQRYSEDPVPLPSET--DGYVAPL	113
Db	1035	PLLSLSLATSNN--NSTVACINRNSCRVKDPAFLQRYSDPTGAVTEDNIDDAFL--	108
Qy	1138	TCSFPEPYVQNPVRPQPSPREGPPAPRPAAGATLERAKTSLPGNGVAVKVPFAGVAV	119
Db	1088	---FVPEYVQ--SVPRRPAQSVQNPVYHQPLHP-----APGRDLHYON--PNSNAV	113
Qy	1198	ENPEYLT-TPQGAAPQPHPPAFSPAFDNLVYWDQ-----DP-----PERGAP	124
Db	1134	GNPEYLYNTAQ-----PTCLSSGFNSPALMIQSHQMSLMDPDYQODFFPKETRYN	118
Qy	1241	STFGKPTPAENPEYGLDVP	1260
Db	1185	GTRKG-PTAENAEYLVAP	1203

```

RESULT 6
epidermal growth factor receptor precursor - chicken
M:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Gallus gallus (chicken)
C>Date: 28-Feb-1996 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
C:Accession: A27720; A00643
R:Laf, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ulrich, A.; Vennart
Mol. Cell. Biol. 8, 1970-1978, 1988
A>Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou
A:Reference number: A27720; MUID:88261272; PMID:3260329
A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A>Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85282822; PMID:2988784
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F:31-654/Domain: extracellular #status predicted <EXT>
F:81-307/Domain: EGF receptor extracellular domain repeat <EE2>
F:397-610/Domain: EGF receptor extracellular domain repeat <EE1>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136,202,280,361,370,422,575,580,615,635/Binding site: carboxylate (Thr) (covalent) #
F:168/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat
Query Match 44.6%; Score 3057.5; DB 1; Length 1223;
Best Local Similarity 48.0%; Pred. No. 2,5e-118;
Matches 625; Conservative 175; Mismatches 352; Indels 151; Gaps 26;
QY 8 RWGLLLALLPPGAA-----STVCTGTDMKRLRLPASPTHLDMLRHLYQGQVYQGNLE 61
DB 13 RGAVALVLLILGVALCSAVEEKVCGQTNKLTOLGVEDHFTSLQRMYNCEVLLSNLE 72
QY 62 LTVLPTNASLSFQDIEGVOGYVLIANQVROYPLQRLIRIVGTQFFENYALATLNDG 121
DB 73 ITVEHNRDLPLKTIQEVAGYVLIANVADVPLNLQIRKNVLYDVSFLAALSNH 132
QY 122 PLNNTPTVTCASPGGIRELQRLSTELIKGVLIQNPOLCYQDTLLMKDIPHKNOAL 181
DB 133 -VNKTQ-----GLRELPMKRLSEILNGVXISNNPKLCNMDDTLVMDIDTSRK-PL 182
QY 182 TLID-TNRSRAHPCSPMKGSRGWSSESDQSLRTYCAAGCA-RCGAPLPTCCHEQ 239
DB 183 TVDPASNLSGCPKCHNCTEDHCKWAGEONCQTLTKVICAOQCSGRGKVPSCCHNQ 242
QY 240 CAAGCTGPKHSDIACILHFNHSGICELHCPALVTYNTDFESMPNDEGRYTFGASCTAC 299
DB 243 CAAGCTGPRSDCLACRKFRRDATCKDCEPLVLYPTTYQMDVAVBEGYSIGALCYAEC 302
QY 300 PNYVLTSDVSSCLVCEPLRHQETIADTGORCEKSKPCARVCYGLGMFNFTVSFFMLRV 359
DB 303 PNYVVTVDHGSVCRRSNTDTEYV-EENGVKCKCKDGLSKVONGIG-----IGELKGI 355
QY 360 PKVSASHLEEFAGCKKIFGSLAFLPSPFDGPASANTAPCYIKANSKFGITELTYLVIS 419

```

```

DB 356 LSIATNIDSFKNCTKINGDVSLIPVAFLODAFTKTLPLDPKKLDVFRVKEISGELLIO 415
QY 420 AMPDLSLDLVFQQLQVIRGRILHNGAYSTLLOGISWLGRLSELESGLALIHNNH 479
DB 416 AMPDNLIDLAPELLEIRKTRKHQGYSLAVNVLKIQSLGSLSEISDGLAIKNNK 475
QY 480 LCFVHTVPMQDLFRNPQALHTANPREDECVGEGLACHQLCARGHGWGPPTQVNCSSQ 539
DB 476 LCVADTNMMSLFTQSGQKTKIIONNKNQCTADRHVCPDLGSDVGCWGGPRHCSGCF 535
QY 540 FLRQCEVBECEVYLCIPREYVARNHCLPHEPCQYQNG---SYTFGEBADQVCACAH 596
DB 536 FSRQKECVKQCNILQGPREFEEDSKCLPCHSECLVQNSTAVYNTQSGDPDHCMAKAF 595
QY 597 KDPFCVACRPSGKPRPLSYMPRIKPEDEGACOPRINCSTSCVDLDDGCAEGRASP 656
DB 596 IDGHCHKACRPAVGLGENDTL-VKRIADANAVQLCHPNCRTGCKRPGLEGCP--NSGK 651
QY 657 LTSIVSAVY-GILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVPLTPSGAMPNQ 715
DB 652 TPSIAGVVGGLLCLVVGGLGIGLYLRR-HIVRKSTLRLLQERELVEPLTPSGEAPNQ 710
QY 716 AQMKILKETELRKXYLGSAGFGTYVKGIMIPRGENYKIPVAIKULENTSPKANEILD 775
DB 711 AHULIKETEFKKVKVLSGAFGVTKGLMIPRGEKVKIPVAIKELREATSPKANEILD 770
QY 776 EAYVMAVGSPIVYSRLIGLCLSTVQLVTLQMLMYGCLLHVEENRGLSODLMMCMOI 835
DB 771 EAYVMAVDNPHVCRLIGLCLSTVQLVTLQMLMYGCLLHVEENRGLSODLMMCMOI 830
QY 836 AKGMSYEDVRLVHRDLAANVLYKSPNNHYKTFGLAALLDIDETIYADGGKVPDKM 895
DB 831 AKGNVYLEERLVRDLAANVLYKTPHNYKTFGLAALLDIDETIYADGGKVPDKM 890
QY 896 ALSEILRRRPTHQSDVSVGVTVWE-LMTFGAKPYDGIIPAREIPDLKEKERLPOPICTI 955
DB 891 ALSEILRRRPTHQSDVSVGVTVWE-LMTFGAKPYDGIIPAREIPDLKEKERLPOPICTI 950
QY 956 DVTIMVYKCMIDSECRPRELIVSEFSSMARDPQRFVYIQ-NEDLGPASPLDSTFYRSL 101
DB 951 DVTIMVYKCMIDSECRPRELIVSEFSSMARDPQRFVYIQ-NEDLGPASPLDSTFYRSL 101
QY 1015 LEDDDMDLVDAREEYLVPOQFGFCPPRAGAGAGMHNRRSSSTSGSGDLTLGEPSEE 107
DB 1011 MEEBDEMDIVDAEYLVPOQFGF-----NSBST----- 103
QY 1075 EAPRSPV-----APSEGAGSDVDDGLGMAAKGLQLPTHPSPLOQYSEDPVLPSE 112
DB 1039 --SRPLLSLSLSTSNNSATNCID-----RNGGCHVREDSPFORSSPTGNFLEE 108
QY 1130 T--DGYVAPLTCSPOEYVNOQDVRPQPSPREBGLPAAAPAGATLERAKTILSGKGVV 118
DB 1089 SIDDGFL-----PABEYVNO-LMKKES-----TAMQONQIY 111
QY 1188 KDVF-----AFGAVENPEYLTPOGGAAPQHPPAFAFPNLTYYMDQ- 123
DB 1120 NMSILRAISKLPMDSRYKSHSANDNPEYL-----NTQSLAKTVFESSFYWIOS 117
QY 1232 -----DPE-----RGAPSTFKGTPTAENPEYLGIDVP 1260
DB 1172 GNRQINLNDNDYQDPLPNETKPNGLIKVPALENPEYLRVAAP 1214

```

```

RESULT 7
A47253
epidermal growth factor receptor, HER4 - human
C:Species: Homo sapiens (man)
C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C:Accession: A47253
R:Plooman, G.D.; Culouscou, J.M.; Whitely, G.S.; Green, J.M.; Carlton, G.W.; Fe
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A>Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the e

```

A:Reference number: A47253; MUID:93189574; PMID:8383326
 A:Accession: A47253
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-1308 <PI0>
 A:Cross-references: GB:I07868; NID:9337359; PIDN:AAB59446.1; PID:9337360
 A:Note: sequence extracted from NCBI backbone (NCBIP:126842)
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; growth factor receptor
 F:716-981/Domain: protein kinase homology <KIN>
 F:724-732/Region: protein kinase ATP-binding motif

Query Match 42.8%; Score 2934.5; DB 2; Length 1308;
 Best Local Similarity 44.7%; Pred. No. 2.9e-113;
 Matches 605; Conservative 188; Mismatches 380; Indels 179; Gaps 29;

9 KGLLLALPFGAA-----STVCTGIDMLRLPASEETHLMDLRLHYGCGVQGNLELY 64
 8 WWSVSLVLAAGTVQPSDSQSVCAAGENTLSSLSLDEQYRALRKYENCEVMGNLEITS 67
 65 LPTNASLFLDIOQVGVVLIHANQVAVPLQRLRIYRGTLFEDNYALAVLDNGPLN 124
 68 IEHRDLSFLRSVRVTVGVVALNQFVLENLRIIGTKLYEDRYALILNTRKDG 127
 125 NTPVTVGASPGGLRELQRLSLTEILKGVLLIQNPQLCYDPTIIMKDIPIKNNQLAVTLI 184
 128 NF-----GLOELGKNTLEILNGGVYDQNKFLCYADTIHMQIIVANPMSNLTIV 178
 185 DTRSRACHPGSPMKSGRSGESEDQSLTRVCAAGC-ARGKGLPTDCCHEQCAAG 243
 179 STNGSSGGRCHKSTG-RCWGPTEHNCQTLRLVCAEQDCGRYGVYDCCCHRECAAG 237
 244 CTGPKHSDCLACLFHNSGICELHCPALVTYNTDTPESMPREGRVTFGASCTACPYNY 303
 238 CSGPKDTCFACMNPENDSGACVTCQPTFYVYNPTTFQLEHNFNKATYGAFCVCKCPHNF 297
 304 LSTDVSGCTVCPHNOEVTADGTQREKSKRCARVCGLGKFNFTVSFWLRPKVS 363
 298 V-VSSSCVRACPSKKEV-BENGKMKKPCCTDIPCACDGI-----TGLMSAQTV 349
 364 ASHLEEFAGCKIFGSLAFLEPSEFGDPASNTAQYKANSKFIETELGYLYISAMP 423
 350 SSNDKFNCTKINNLIFLVTHGDPYNAIEAIDEXKLVNFTVETIGFLNIGSWP 409
 424 SLPLSVFQNLQVTRGLHNGANSLTQIGISWLGRLSIREGSLALIHNTLCFV 483
 410 NMTDFSVSNLVTIGRVLVSGLSILLIKQOGITSLOFOSLKEISAGNIYITDNSLCY 469
 484 HTVPWDCLEFRPHOALHTANRPEDECGEGLACHOCACAGHGMCPPTOCVACSGFLG 543
 470 HTINMTLFTINQRIYIRDNKRAKENTCAEGMVCNHLCCSDCKGPPDCLSCRFPSRG 529
 544 QECVECEVLOGLPREYVNAHCLPCHPECOP-ONGSVTCFGEADOCVACAHYKDPFC 602
 530 RICESCNLYGGEFEFENGSIQVECDPQCEKXEDGLTCHGROPNCTKCSHFKGPNC 589
 603 VARCPGVKPLSLVMPIMKPFDEGACQPCINCTHSCVLLDKG-----PAE 651
 590 VEKCPDGLQGANRF-IFKYADPRECHPCHPNTQCGNPTSHDCIYFMTGHTLPH 647
 652 QRASPLTSIVAVV-GILLVVVLGVFGLIKRQCKIRKYTRRLDQETELVEPLTPSG 710
 648 AR-TPL--IAAGVIGGFLIVTIGLFAVYVRSIK-KBALRRL-ETELVEPLTPSG 702
 711 AMPNOQWELIKELKRVKYLGSAGACTYKGMVGDGNVKAIPVAILKRLNTSPKAN 770
 703 TAPNOQRLIKELKRVKYLGSAGACTYKGMVGEVTVKIPVAILKRLNTSPKAN 762
 771 KEIDEAAYVAGVSPVYRLLGLTSTVQLVTLMPYGCCLDHYRENGRGLSGDOLL 830
 763 VEFMEALIMASMDHPHLVRLGVCSPTIQLVTLMPHCCLEVEYHEKDNIGSGLLN 822
 831 WCMQIAKMSYLEVRLVHNDLAARNVLVSPNHYKITDGLARLLDIDETEVHADGKV 890

DB 923 WCVQIAKMSYLEERLVRDLAARNVLVSPNHYKITDGLARLLDIDETEVHADGKM 88;
 QY 891 PIKMALESILRRRPTHOSDVWSGYTVWELMTGAPYDGIAPAREIPDLLEKERLP 950
 DB 883 PIKMALESILRRRPTHOSDVWSGYTVWELMTGAPYDGIAPAREIPDLLEKERLP 94;
 QY 951 PICTIDVYMWKCMKIDSECRPPREIVSEFSMAWDPQRFVYIQNEI-LGPASPLDST 100
 DB 943 PICTIDVYMWKCMKIDSECRPPREIVSEFSMAWDPQRFVYIQNEI-LGPASPLDST 100
 QY 1010 FYRSLLEDMDGDLVDAEYVVPQGFPCPPAPACAGMHHRRSSSTSGGDLTLG 106
 DB 1003 FQNLIDEDLEDMDWDAEYVLP-QAENIIPP-----ITSRRLIDSNRS-----EIGH 106
 QY 1070 EPSEEARPS-----PLAP-SEGASDVFPDGLCMGA 110
 DB 1051 SFPAYTSMGNQFVYRDCGFAAGVSVFAPRTSTIIPAPVAGATAIFFDSCNGT 111
 QY 1101 AKGLQSLPTHSPFLQYSEDPYPLPS-----ETDGYVAPLTGSPQRYNQPVRP 116
 DB 1111 LRKYAPHVQDSSSTQYSAQPTVFAERSPRGELDEGYMTPEYRDKPKQELNPVE 116
 QY 1154 QPPSPREGLPAPAPAGATLERAKTILSPKNGVYQVFAFGAVENPEYLTPQGAAP 121
 DB 1168 -----ENPFVSR-----KNGDLQ-----ALDNEFYHNASNG----- 116
 QY 1214 HPPPA-----FSPAFNLVYWDQDPERGA- 123
 DB 1195 -PPRADEYVNEPLVNTFANTLCAEYLNKNNILSMPEKAKAFDNDPNHSLPRASTL 125
 QY 1239 -PSTFKGTPT-----AENPEYL 1255
 DB 1254 QHPDLYGYSIKYFYKNGRIRIVANPEYL 1285

RESULT 8
 506142
 protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish
 N.Alternate names: epidermal growth factor receptor homolog; kinase-related tr
 C.Species: Xiphophorus maculatus (southern platyfish)
 C.Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 18-Feb-2000
 C.Accession: S06142; S13809
 R.Witcomb; J. Adam; D. Maltzschek; B. Maueleler; W. Raulf; F. Telling; A
 Nature 341, 415-421, 1998
 A>Title: Novel putative receptor tyrosine kinase encoded by the melanoma-induc
 A.Reference number: S06142; MUID:90015140; PMID:27977166
 A:Accession: S06142
 A:Molecule type: DNA
 A:Residues: 1-1166 <MT>
 A:Cross-references: EMBL:X16891; NID:965290; PIDN:CAA34770.1; PID:965291
 R.Adam; D. Maueleler; W. Scharf; M.
 Oncogene 6, 73-80, 1991
 A>Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in
 A.Reference number: S13807; MUID:9115882; PMID:1846957
 A:Accession: S13809
 A:Status: preliminary; translation not shown
 A:Molecule type: DNA
 A:Residues: 821-1025 'N', 1027-1098 'A', 1100-1166 <ADA>
 A:Cross-references: EMBL:X66319; NID:965284; PIDN:CAA39763.1; PID:965285
 C:Genetics:
 A:Gene: mrk
 A:Map position: Y
 A:Insertions: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane pro
 F:1-25/Domain: signal sequence #status predicted <Sig>
 F:26-1166/Product: kinase-related transforming protein (Tu) #status predicted
 F:707-972/Domain: protein kinase homology <KIN>
 F:715-723/Region: protein kinase ATP-binding motif

Query Match 38.5%; Score 2639; DB 1; Length 1166;
 Best Local Similarity 44.5%; Pred. No. 3.4e-101;

Matches		568; Conservative	173; Mismatches	384; Indels	152; Gaps	32;
Qy	4	AALCRMGILLALPPGAASST-----OVCGTDMKLRLPASPETHLDMLRHLYOGQGVQGN	59			
Db	8	AAALQ--LILVLISIRCCSTDPDRKVCQSTNQMIM--LDNHTLKMKKMSGGNVLEN	62			
Qy	60	LELYLFTNASISFLQDIOEVQGVYLLAHNOVROVPLORLRIVRGTLFEDNVALAVLDN	119			
Db	63	LEITYQENQDLSFIQSIQEVQGVYLLAHNEVSTIPLVNLRLINGQVLYEENFTLVMSN	122			
Qy	120	GDPLNNTPTVTGASGGRLBELQRLSTELIKGVILQINPOLCVQDITLMKDITHKMQL	179			
Db	123	VQK-NPSSP--DIVQVQKQQLQSLNLTILISGGVAKSHNPLLCVETIIMMDIVIDKTSNP	179			
Qy	180	ALLTIDNRSRACHPCSPMKSCRCMGSESSDCSLRTVCAGGC-ARCKGPLPTDCHE	238			
Db	180	TNNLIPLHAFERQCGQDHGCVNCSWAPRGHCCKFTLLCAEQCNRCRGPKRIDCNE	239			
Qy	239	QCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESPNPEGKRTGASCVTA	298			
Db	240	HCAGGCTGPRATDCLACRDFNDGCKQTCPPKRIYDIVSHQVVDNPNIKYTFGAAVCYE	299			
Qy	299	CPYNYLSTPVGSCITLVCPHNOEYTAEDGTORCEKSCAPACVCGMFN-NFTVFWL	357			
Db	300	CPENYVTE-GACVASCAGMLEVD-ENGKSCCKPCGCGVCVKCDGIGISLNTIA---	354			
Qy	358	RVPKVSASHLEEFACCKIFQSLAFSPESFDGP--ASNTAPQYIKANSKFIGITELTG	414			
Db	355	---VNSTNIRSFNSCTKINGDIILNRSFEGDPRHYKIGTMDPEHL---WNLTVEKITG	407			
Qy	415	YLISAMPDSDLPDLSPONLOVIRGIRLHNGAYS-LTLQIGISMLRLSRRELGSGLAL	473			
Db	408	YLIVMMPEBNMISLSVFOVLEIRKRTTFSGFSFVAVQVRHLQWLRLSRKEVASANVI	467			
Qy	474	IHNHTLQFVHTVPMDQLFNPHQALLHTANRPDECEGELACHQUCASGHCGWGPQTQ	533			
Db	468	LKNTLQLRVANTINMRLLFRSEDOGIEVDART-----ENQTCNNCSGDGCM-PEPTM	519			
Qy	534	CYVCSQFLRQGECEECVQLQGLPREYVNAHCLPCPECCOPQNGSYTCGPEADQCVAC	593			
Db	520	CYSCILHVRGRCAVASCNLLQGEPEACVDRCVQCHQCECLVQDLSITCYGPGRANCSK	579			
Qy	594	AHYKDPFCVACPCSGVNPDLSPYIMKFPDEBACQPCPCINCHSCVDDDDKGCPEQR	653			
Db	580	AHFQDQPCQIRCRPHGILGGDGL--IMKYADKMQCQPCCHNCOCSSGSGLSGCRD-1	637			
Qy	654	ASPLTISAVVGLILVVLGVVEGILIKRPOQKIRKYTWRELLIQETELVEPLTPSGAMP	713			
Db	638	VSHSLAVGLVSGLLITVIVALLTVLLRRRIK-RKRTIRCLLQEKELVEPLTPSGAP	696			
Qy	714	NOACWRILKETELRKVKYLGSAGFTVYKGIWPDGENVKIPVAKYLRNTSPKANKET	773			
Db	697	NOAFRIKETEFKCDRLVSGAGFTVYKGLMNDGENIRIPVAKYLRNTSPKANKOEY	756			
Qy	774	LDEAVYVAGVSPYVSRLLGICLTSTVQVLTQMLPYGCLLDHYENRGLSGSODLNMCM	833			
Db	757	LDEAVYVAVSVHPVCHRLGICLTSAVOLVQMLPYGCLLDYVHQHERICQGWLLNMCV	816			
Qy	834	QIAKMSYLEDVRLVHPRLAARVNVKSPNNVKITTDGLARLDDIDETEAAGKVPK	893			
Db	817	QIAKMNLEERHVLHRLAARNVLLKXPNVKITTDGLSLTLADKEKQAOAGKVPK	876			
Qy	894	WMALESILRRFTQSDVMSYGVTVWELMTFGAKPYGDIIPAREIPDLLKGERLPPQPC	953			
Db	877	WMALESILQWYTHQSDVMSYGVTVWELMTFGSKPYGDIIPAKELASTLENGERLPPQPC	936			
Qy	954	TIDVYMIWKCMMDISEGRFRFRELVEFSNMADPQRFVITQMEDIGRASPILDSTFYNS	1013			
Db	937	TIEVMIILKCMMDIPSSRFRFRELVEFSQMAPDPRKLYIQO--NPLSLDRKLFGR	993			
Qy	1014	LLEDDMGDLVDAEYLVPOCGFCPPAPAGAGMVHRRSSSTRGCGDLTLGSEPS	1073			
Db	994	LLSDD--DVVDADBYLLPKRI-----NRGS-----	1019			

Qy	1074	EEAPRPLAPSEGAGSDVFDGDLGMAKGLQSLPTHDPSPQLQRYSEDPRV-P-PSRTG	11;
Db	1020	-----EPCLPRTGH-----PVRENSTTLRNISDPIONALEKLDG	10;
Qy	1133	YVAPLTCSPQPEEYVNPDPVRPQ-----PSPRE-----GPLP-AARPGATLERAKTL	11;
Db	1055	H-----EYVNPQSESTSSRLSDIYNPNYEDLTDGMCQPVSLSSQEAETNFSRPREV	11;
Qy	1180	SPGKGVKDVFAFGAAGVENEPEYLTPOGGAAPQHPRPAPSPARDNLVYMDQPPREGAR	12;
Db	1105	NTNQNLSL--PLVSSGMDDDPY---QAG-----YQAF-----LPQTGAL	11;
Qy	1240	PSTFKGTPTAENPEYLG	1256
Db	1140	TQNGMFLPRAENLEYLG	1156

RESULT 9
 A36223
 Kinase-related transforming protein (erbB3) (EC 2.7.1.1) precursor - human
 C.Species: Homo sapiens (man)
 C.Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
 C.Accession: A36223; 159164
 R.Kraus, M.H.; Issing, W.; Miki, T.; Pospescu, N.C.; Aaronson, S.A.
 Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
 A.Title: Isolation and characterization of ERBB3, a third member of the ERBB/
 A.Reference number: A36223; MUID:90083234; PMID:2687875
 A.Accession: A36223
 A.Status: preliminary
 A.Molecule type: mRNA
 A.Residues: 1-1342 <KDA>
 A.Cross-References: GB:M29366
 R.Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.;
 Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
 A.Title: Molecular cloning and expression of another epidermal growth factor
 A.Reference number: 159164; MUID:90311312; PMID:2164210
 A.Accession: 159164
 A.Status: preliminary; translated from GB/EMBL/DBJ
 A.Molecule type: mRNA
 A.Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>
 A.Cross-References: GB:M4309; NID:g183990; PIDN:AAA35979.1; PID:g306841
 C.Genetics:
 A.Gene: GDB:ERBB3; HER3
 A.Cross-References: GDB:119880; OMIM:190151
 A.Map position: 12q13-12q13
 C.Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kit
 C.Keywords: ATP; phosphotransferase
 F:707-972/Domain: protein kinase homology <KIN>
 F:715-723/Region: protein kinase ATP-binding motif

Query Match		34.8%; Score 2382.5; DB 2; Length 1342;	
Best Local Similarity		39.9%; Pred. No. 1.38-90;	
Matches		525; Conservative 196; Mismatches 461; Indels 135; Gaps 32	
Qy	10	GILLALLPQGA--STQVCTGDMKLRLPASPEHLDMLRHLYOGQGVQGNLELTLP	67
Db	11	GULSLAAGSEVNSQAVCPETLNGLSVTGDAEQVQTLKYLERCGVYVNGNLEIVLTGH	70
Qy	68	NASLFLQDIOEVQGVYLLAHNOVROVPLORLRIVRGTLFEDNVALAVLDNCDPLNNT	127
Db	71	NADSLQWIREVYGVYLVANNESTLPLNLRVRYGTQYDDKFAIFVW-----LNYNT	125
Qy	128	PVTGASPGGLRELOLRSTELIKGVILQINPOLCYQDITLMKDITHKMQLALTITDN	187
Db	126	-----NSHARQLRLTQLTIELSGVYIEKDKRICHWDITDMWDIVARD--AETVYVD	178
Qy	188	RSRACHPCSPMKSCRCMGSESSDCSLRTVCAGGC-ARCKGPLPTDCHEQCAAGCTG	246
Db	179	NGRSCPRHEVCCK-RCKMGSESECCQLTLITICAPQCNHGCFEPNRYQCHDCAGCGS	237
Qy	247	PKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGKRTFGASCVTACPVYLT	306


```

Db      238 PDDTDFACGHEFNDGACVPRCPQLVYNNKLTFTGLEPNHTKYQYGVASCPHNFY-V 296
Qy      307 DVGSCITLVGRLNNGEYTAEDGTQRCCKSKPCARVCYGLGEMNNTFVWLRYVYSASH 366
Db      297 DQTSVRAQCPDPMEDV-KNGLMCEPCGGLCPKACEGSGSSRFQT-----VDSNN 347
Qy      367 LEEFAGCKKIFGSLAFPESEFDOPASNTAPQYIYANSKFIGITLTGYLYISAMPDLP 426
Db      348 IDGFVACTKLGNLDPLITGLNDPMHKI.PALDPELNVFRVREITGYLNIQSPRMHM 407
Qy      427 DLSVFONLOVIRRIIHNAGYS-LTQIGIGISWLGIRSRRELGGSLIHNTLCVHT 485
Db      408 NFSVFSNLTTTIGRSLYNRGFSLLMKNLNVSLGRSLKEISAGRVIYISARQCYHNS 467
Qy      486 VPMDQTFRNPHQALLHTA-NREDECVGEGLAHQICARHCWGEPTQVCNCSQFLRGQ 544
Db      468 LNWTKYLRGTEERLDIKNNRRPCDVAAEGKYCDPLCSGGCGWGPQGLSCRYNSGG 527
Qy      545 ECVEECRVLGRLPREVYNARHCLPCHPECOPONGSVTCGPPADQCVAAHYKQPPCVA 604
Db      528 VCETHCNFLNGEPRERFAHEAFCSCHPECOPMEGTATCGSGSDTCACCAHFRDPPHCVS 587
Qy      605 RCPGKYKPLDLYMPIKPFDEGACOPDINGTHSCVDLDDKCPABQRA-----SPLTSI 660
Db      588 SCHGVUG--AKGPIYKYPDVQNECRPCHENCTQCKGPELODCLGQTLVLIGKTHLTA 645
Qy      661 VSAVVGILLVVLGVVFGILLKRRQOKR-KYTNRRLLQETELVEPLTPSGAMPQOQMR 719
Db      646 LTVIAG--LVVIFPMIGGTFELWGRRIQNKRAMRYLERGESIEPLDPS-EKANKYLAR 702
Qy      720 ILKETELRKVKYLGSGAFGYVYKGIWIPDEGVKIPVALKYLRNTSKRANKELIDEXYV 779
Db      703 IFKETELRKLVKVGSGVFGTVHKGWIPGESIKIPVCKIYLEDGSGQSOAVLTDHMLA 762
Qy      780 MAGVSPVYSRLIGITLTSTVQLTVMYGCILDHVENRGLSQDLTMMQOIAKAM 839
Db      763 IGLDHAHIVRLIGLPGSSLOLVQYPLGSLDHVQHRLGALGRLMMQOIAKAM 822
Qy      840 SYLEDVRLVRLDAANNVLYKSPBNYKIDFGALALIDETETNADGQKVPYKMALES 899
Db      823 YLLEHGHWHRNLAAANNVLKSPQOVADFGVADLLPPDDKQLYSBAKPIKMALES 882
Qy      900 ILRRFTHOSDVWSYGVTVWELMTFGAKPYDGI.PAREIPDLEKGERLPOPICTIDVYM 959
Db      883 IHFGKXTHOSDVWSYGVTVWELMTFGAEYAGLRILAEVLDLEKBERLAQOICTIDVYM 942
Qy      960 IMYKCMWIDSECRPRELVESEFMYARDPQRFVVIQNEULGPA---SPLDSTFYSLLE 1016
Db      943 VMYKCMWIDENIRPTEKELANETRMARDPPRYLVIKRES-GPGIAPGPEPHGLTNKYLE 1001
Qy      1017 DDDMGDLVAEEYLPOGCFPCDPAPAGAGWVHHRHSSSTRSGGDLTLGLEP-SEE 1075
Db      1002 EVELPELDDLDLEED-----MLATTLTIGSALSLPVGTLNPR 1042
Qy      1076 APPSPAPSEAGSDVFDGLGMAKGLQSLPTH-DSPLORYSEDPVLVLP-----S 1128
Db      1043 GSGSLTSPSSGY-MPMNQNLGSCQESAVSSSERCPRVSLH-----PMRPGCLASE 1095
Qy      1129 ETDGYYA-----PLTCSQPR-----YNNQPRVQPPSPREGP----- 1162
Db      1096 SSEGHVTGSEAEIQEVSMCRSRSRSPRPGDSAYHSQRHSLTPTPLSPGLEED 1155
Qy      1163 -----LPAARPAQATLERAKTLP-S-GXNGV-----KDVFAFGAVNPEYLPFGGAA 1210
Db      1156 VNGXVWPDTHLKGTPSREGTLSSVGLSGLTEBEDD-----EYYNNRRRRHS 1207
Qy      1211 POHPPPAFSPADNLVYWD-----QDPPEGAPSTFKGPTLVENDEYL 1255
Db      1208 P-PHPRPSSLEELGYEYMDVGSDSLASIGTQSCPLHPVIMPPTAGTTPDEDEYEM 1263

```

```

epidermal growth factor receptor homolog precursor - rat
N/Alternate names: ErbB3 protein; HER3 protein
C/Species: Rattus norvegicus (Norway rat)
C/Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C/Accession: J04387
R/Hellier, N.J.; Kim, H.H.; Greaves, C.H.; Sterke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A>Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinan
A/Reference number: J04387; MUID:96096535; PMID:8522190
A/Accession: J04387
A/Molecule type: mRNA
A/Residues: 1-1339 <HEL>
A/Cross-references: GB:U29339; NID:9915389; PID:9915390
A/Experimental source: liver
A/Note: The authors translated the codon AAC for residue 369 as Thr and GTT f
C/Comment: This protein is a functional heregulin receptor that transduces si
C/Genetic:
A/Genes: ErbB3
C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein ki
C/Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane
F/1-19/Domain: signal sequence #status predicted <SIG>
F/20-139/Product: epidermal growth factor homolog #status predicted <MAT>
F/640-659/Domain: transmembrane #status predicted <TM>
F/705-970/Domain: protein kinase homology <KIN>
F/713-721/Region: protein kinase ATP-binding motif
F/939,1051,1156,1196,1219,1257,1259,1273,1286,1325/Binding site: phospho
Query Match 33.5%; Score 2294.5; DB 2; Length 1339;
Best Local Similarity 40.0%; Pred. No. 5,1e-87;
Matches 515; Conservative 176; Mismatches 437; Indels 161; Gaps 3
Qy      3 LAALCRWGLLALLPRGA---STQVCTGTDMKRLPASPETHLDMRLHYQSGQVYVGN 59
Db      7 LQVLC---FLTLARGSBMGNSQAVCCGTLNGLSVTDADNQVQYLYKLEKEVWGN 62
Qy      60 LELTYIPTASISFLODQVQGYTLIHNQVQVPLQRLATVGTQVFEENYLAVIDN 11
Db      63 LELTVLGHNAIDSLFQWIREVYAVLVAMNESVPLFNLNVGTVQYDGKFAIPVM-- 12
Qy      120 GDPLNNTFVTGASPGGLRELQRLSTELIKGVLLIQNPQCYODTILMDIFHKNNQL 17
Db      121 ---LNVNT---NSHALRQLKFTGLTEILLSGGVYIEKNDLCHMDITDWRDIYVR-- 17
Qy      180 ALTLIDTNRSAQHCHSPCKGSRGWSSSEDCQSLRTTVAAGC-ARCKPRLPTDCHE 23
Db      171 GAEIVKNNGANCPCHVEYCKG-KCMGGGPDQQLITLTICAPQCNCGFCGPNPNQCHD 22
Qy      239 OCAAGCTGKSHDCLACHFNHSGICEHCPALVTVYNDTFESMPNPRGRTFGASCYTA 29
Db      230 ECAAGCGSPQDIDCFACRRFNDSGACVRCPEPLVYNNLTQLERNPTTKYQYGVVYS 28
Qy      299 CPYNYLSTDVGCTLVCPRLHNOEYTAEDGTQRCCKSKPCARVCYGLGMFNNFTVSEFLR 351
Db      290 CPHNFY-VQTFPVARACPPDKMEVD-KHGLMKCEPCGGLCPKACEGSGSS-----R 33;
Qy      359 VKVYSALHEEPAAGKKTIFGSLAFPESEFDOPASNTAPQYIYANSKFIGITLTGYLY 411
Db      340 YQTVDSNNDGVNCTKLIGNDPLITGLNDPMHKI.PALDPELNVFRVREITGYLNI 391
Qy      419 SAMPSLPLDSVFNLOVIRRIIHNAGYS-LTQIGIGISWLGIRSRRELGGSLIHNTLHN 477
Db      400 QSMPPHMHNFVSFSLTTIGRSLYNRGFSLLMKNLNVSLGRSLKEISAGRVIYISAN 455
Qy      478 TILCFVTHYPMQLFRNHQALLHTA-NREDECVGEGLAHQICARHCWGEPTQVCN 536
Db      460 QQLCYHNSLMTWRLIRGSEERLDIKYDRPLGEBLAEKVCYDPLCSGGCGWGPQGLS 515
Qy      537 GQFLRGQCEVEECVTLGRLPREVYNARHCLPCHPECOPONGSVTCGPPADQCVAAHY 596
Db      526 GKNYSREGVCTHGNFLGGEPRERFAHEAFCSCHPECOPMEGTATCGSGSDTCACCAH 575
Qy      597 KDPFCVACRPSGVKPLDLYMPIKPFDEGACOPDINGTHSC--VDLDDKCPABQRA 654

```

```

Db      580 RDGPHCVNSCPHGLC--AKGPIYKYPDAQNRCRPHENCTGCGNPELODCLGAEVLM 637
QY      655 SPTISAVAVGLLVVGLVGVGLIKRROOKIR-KYTMRLLOETLEVEPLTPSGAMP 713
Db      638 SKPHVIAVTVG--LAVIIMILGSGFLYKRGRIQKAMKMYLERGESIEPLDPS-EKA 694
QY      714 NQAKMRLKETELRKVKYLGSGAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPANKKEI 773
Db      695 NKYLARIFKETELRKVKYLGSGVGFVHKGIWIPGESIKIPVCIKVIEDKSGRQSPQAV 754
QY      774 LDEAYVAGVGSFYVRLLIGLCTSTVOLVTLQMPGCLLDHVRNRRGLSGDILNMC 833
Db      755 TDMLAVGSLDAHVRLLGLCGSSQLVLYVLPGLSLDHVKQKRETLQGLLNMGV 814
QY      834 QIAKMSYLEDVLRHDLAARNVLYKSPNHVKTDFGLARLLDIDETEHADGGKVPK 893
Db      815 QIAKMYLEESHWVRDLARVMYKSPSYQVADDFVADLLPDDKQLLHSEAKTIX 874
QY      894 WMALESLRRRTFHOSDWSYGVTVWELMTFGAKRPDGIIPAREIDDLLEKGRLLPQPIIC 953
Db      875 WMALESLHFGKYTHOSDWSYGVTVWELMTFGAEPYAGRLAEIDDLLEKGRLLAQPIIC 934
QY      954 TIDVIMVAKMWIDSECRPFRELVEFSRMAPDPQRFVVIQNEIDLGPASPLDSTFYRS 1013
Db      935 TIDVYMWVYKMWIDENIRPTFRELANETPRADPRYLVIKRS--GRGP--PAAPS 991
QY      1014 LEEDDMKGLVDAEYLVVPOQGFCDPRAPGAGVHHRHRSSTRSGGDLTGLEPSE 1073
Db      992 VLTTEL-----QEALPEL-----DLDLLEAES 1017
QY      1074 E-----EAPRSLAPSEG-----AGSDVPDGLGMAKALQ 1105
Db      1018 EGLATLSGSLSPCTGLTRPRSGQLSPSSGVPYPMNOSLGEACLDASVAGCEQFSR 1077
QY      1106 SLTPHPSPLQRYSEDPVLPVLPSETDGYV---APL-----TC-----SPOE---Y 1145
Db      1078 PLSLH-PIPRGR-----PASSESGHYTGSEAEIQEKVSVCFSRSRSPRPGRDGA 1129
QY      1146 VNQPDVPRQPPSPREGP-----LPARPGATLEPAKTLSP-QKRGV----- 1187
Db      1130 HSGRHSLLPTVPLSPGGLLEBEDGNGYVMPDTHLRGASSREGTLSSVGLSTVEBED 1189
QY      1188 KDVFAGFCAVENPEYLLPQGAAPQPPHP 1216
Db      1190 ED-----EEYEVNKRKRQSP-PRPP 1209

RESULT 11
TVFVLV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N:contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C/Species: avian leukosis virus, ALV
C/Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C/Accession: B00643; A00643
R/Nielsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A/Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and p12
A/Reference number: A00643; MUID:85228222; PMID:2988784
A/Accession: B00643
A/Molecule type: mRNA
A/Residues: 1-698 <NLS>
A/Cross-references: GB:M10066; GB:M13881; NID:g211749; PIDN:AAA48763.1; PID:g211750
A/Note: In Genbank entry CHKEBHF, release 109.0, the source is designated as Gallus gallus
C/Comment: This protein is synthesized as a gag-env-erbB protein.
C/Genetics:
A:Gene: gag-env-erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
P; 1-6/Protein: gag protein (fragment) #status predicted <GAG>
F; 7-59/Protein: env protein (fragment) #status predicted <ENV>
F; 60-698/Protein: protein-tyrosine kinase erbB #status predicted <ERB>
F; 194-459/Domain: protein kinase homology <KIN>
F; 202-210/Region: protein kinase ATP-binding motif

```

```

F;229/Active site: Lys #status predicted

Query Match      25 %; Score 1766.5; DB 1; Length 698;
Best Local Similarity 52.2%; Pred. No. 12-65;
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18

QY      584 GREADQVCAHYKDPFVCARCPGSKDLSYPMIWKFPDEGACQCPRLCTHSCVDL 643
Db      60 GP--DHQMCAPHIDGPHCVKACPAVLGENDTL-VMKYADANAACQCLPCHCTGCKGP 116
QY      644 DKGCCAEQASPLTISAVV-GILVVYLVGVGLIKRQOKIRKYTMRLLOETLE 702
Db      117 GLEGC--NGSKTPSIAGVVGGLLVGLVGLVLR--HIVRKTLRLLQEBEL 172
QY      703 VEDLTPSGAMPNQAQRILKETELRKVKYLGSGAFGTVYKGIWIPDGENVKIPVAIKVLR 762
Db      173 VEDLTPSGAMPNQAQRILKETELRKVKYLGSGAFGTVYKGIWIPDGENVKIPVAIKVLR 232
QY      763 ENTSPYANKKEIDEAIVMAVGSFVYSRLGICLTSTVOLVTLQMPYGCCLDHVRNRR 822
Db      233 EATSPYANKKEIDEAIVMASVNPVHCRLLGICLTSTVOLVTLQMPYGCCLDIYREHKN 292
QY      823 LGSQDLNMCQIAKMSYLEDVLRHDLAARNVLYKSPNHVKTDFGLARLLDIDETE 882
Db      293 IGSQYLLNMCQIAKMSYLEDVLRHDLAARNVLYKSPNHVKTDFGLARLLDIDETE 352
QY      883 YHAGGKVPKIMMALESILRRRTFHOSDWSYGVTVWELMTFGAKRPDGIIPAREIDDLLE 942
Db      353 YHAGGKVPKIMMALESILRRRTFHOSDWSYGVTVWELMTFGAKRPDGIIPAREIDDLLE 412
QY      943 KGRRLQPPICITIDVYMWVAKMWIDSECRPFRELVEFSRMAPDPQRFVVIQNEIDL 100
Db      413 KGRRLQPPICITIDVYMWVAKMWIDSECRPFRELVEFSRMAPDPQRFVVIQNEIDL 472
QY      1002 PASPLDSTFYRSLEDDDMKGLVDAEYLVVPOQGFCDPRAPGAGVHHRHRSSTRSG 106
Db      473 LPSPTSKRYRLMEBEDMEDLVDAEYLVVPOQGF-----NSPT--- 513
QY      1062 CGDULTGLEPSEEAERSL-----APSEGAGDVFDGLGMAKALQSLPTHDPSPLQ 111
Db      514 -----SRPLSLSATSNNSATCID-----RNGQGHVREDSTFVQ 550
QY      1117 RYSEDPVLPSET--DGVAVPLTCSQPEVYNQPDVAPPPSPREGFLPAARPGATLE 117
Db      551 RYSEDPVLPSET--DGVAVPLTCSQPEVYNQPDVAPPPSPREGFLPAARPGATLE 585
QY      585 RYSEDPVLPSET--DGVAVPLTCSQPEVYNQPDVAPPPSPREGFLPAARPGATLE 633
Db      1220 SPAEDNLYWQ-----DPE-----RGAPPSFFKGTPTAENPEYGLDVP 1260
QY      634 KTVFSSPYWISGNHQINLDPDQDPLPNETKNGLLKVPAPENPEYLRVAP 689

RESULT 12
TVYVH
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (src)
C/Species: avian erythroblastosis virus
C/Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C/Accession: A00644; A38022
R/Kamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A/Title: The erbB gene of avian erythroblastosis virus is a member of the src g
A/Reference number: A00644; MUID:84026539; PMID:6313229
A/Accession: A00644
A/Molecule type: DNA
A/Residues: 1-604 <YAM>
A/Cross-references: GB:K01216; NID:g209676; PIDN:AAA42400.1; PID:g209678
R/Dehnbire, B.; Henry, C.; Betalasa, M.; Bisette, G.; Claverie, J.M.; Saule, S.;
Science 224, 1456-1459, 1984
A/Title: Sequencing the erbB gene of avian erythroblastosis virus reveals a new
A/Reference number: A38022; MUID:84232957; PMID:6328658

```


A/Accession: A38022
 A/Molecule type: DNA
 A/Residues: 128, 'W', 30-139, 'F', 141-145, 'V', 147-152 <DEB>
 A/Cross-references: GB:K02006
 A/Genetics:
 A/Gene: erbB
 C/Superfamily: epidermal growth factor receptor; protein kinase homology
 C/Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
 F130-395/Domain: protein kinase homology <KIN>
 F138-146/Region: protein kinase ATP-binding motif
 F165/Active site: Lys #status predicted

Query Match 24.9%; Score 1703; DB 1; Length 604;

Best Local Similarity 52.2%; Pred. No. 4.2e-63;
 Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

```

QY 593 CAHYKDPFCVACRPSGVKPDLSYPIWKPRDEGACOPRINCTHSCVDLDKGGPAAO 652
DB 3 CAHFIDGPHCVKACRPAVLGENDTL-VKRYADANAVCOLCHPCTRCCKGPGLEGP--- 58
QY 653 RASPLTSIVSAVV-GILLVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGA 711
DB 59 NGSKTPSIAGVVGGLCLVVGIGLGLYLRH-HIVKRTLRLLQERELVEPLTPSGE 117
QY 712 MPNOAKMILKETLRKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANK 771
DB 118 APNOAHLRLKETEFKVKVKGSAFGTVYKGIWIPDGENVKIPVAIKELRENTSPKANK 177
QY 772 EILDEAYVAVGVSFVYSRLIGICTSTVOLVQLMYPGCLLDHVENRQRLGASODLNM 831
DB 178 EILDEAYVAVGVSFVYSRLIGICTSTVOLVQLMYPGCLLDHVENRQRLGASODLNM 237
QY 832 CMQIAKGNLYLEDVRLVHRDLAANVLVKSPPNHVKITDFGLARLDIDETEYHAGSKVP 891
DB 238 CVOIAKGNLYLEDVRLVHRDLAANVLVKSPPNHVKITDFGLARLDIDETEYHAGSKVP 297
QY 892 IKMALESILRRRTTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPORP 951
DB 298 IKMALESILRRRTTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPORP 357
QY 952 ICTIDVYIMVYKCMWIDSECRPRELVSFESRVARDPORFVYIQ-NEDLGASPLDSTF 1010
DB 358 ICTIDVYIMVYKCMWIDSECRPRELVSFESRVARDPORFVYIQ-NEDLGASPLDSTF 417
QY 1011 YRSLLEDODMGDLVDAEYVLVPOGFCPPRPARAGAGMHHRRSSSTSGGDLTLGLE 1070
DB 418 YRSLLEDODMGDLVDAEYVLVPOGFCPPRPARAGAGMHHRRSSSTSGGDLTLGLE 449
QY 1071 PSBEAPRSPL-----APSEGAGSDVFDGDLGMAKAGLSLPTHPSPLORYSEDPVP 1125
DB 450 -----SRTPLLSLSATSNNSATNCID-----RNGQGHVREDSDFYQRYSSDPTGN 495
QY 1126 LPSET--DGYVAFITCSPQPEYVQVPRQPPSPRGGPLPAARPAAGATLERAKTISPK 1183
DB 496 FLESIDDGFL-----PAPRYVNO-LMPKKPSTAN----- 524
QY 1184 NGVYKDVFAF-----GAVENPEYLPQGAAPORPPAPSPAPD 1224
DB 525 --VONQIYNFISLAIKSLPMDSRYGNSHSTAVDNPEYL-----NNQSLATLATIVE 574
QY 1225 NLVYWDQPPRGAAPSTFKGTPTAENPEY 1254
DB 575 SSPYWIQSGNHQ-----INLDNPDY 594

```

RESULT 13

S35745
 protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
 C/Species: avian erythroblastosis virus
 C/Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
 C/Accession: S35745
 R/Vennstrom, B.
 submitted to the EMBL Data Library, March 1993

A/Reference number: S35743
 A/Accession: S35745
 A/Molecule type: DNA
 A/Residues: 1-544 <VEN>
 A/Cross-references: EMBL:X12707
 A/Genetics:
 A/Gene: erbB
 C/Superfamily: epidermal growth factor receptor; protein kinase homology
 C/Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-
 F135-400/Domain: protein kinase homology <KIN>
 F143-151/Region: protein kinase ATP-binding motif
 F170/Active site: Lys #status predicted

Query Match 24.0%; Score 1647; DB 2; Length 544;

Best Local Similarity 54.9%; Pred. No. 7.4e-61;
 Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15

```

QY 584 GPEADQVCACVAKDPFCVACRPSGVKPDLSYPIWKPRDEGACOPRINCTHSCVDL 643
DB 1 GP--DHCMCAHFIDGPHCVKACRPAVLGENDTL-VKRYADANAVCOLCHPCTRCCKGPG 57
QY 644 DDKCPAQPASPLTSIVSAVV-GILLVVLGVVFGILIKRQOKIRKYMRLLOETEL 702
DB 58 GLEBCP---NGSKTPSIAGVVGGLCLVVGIGLGLYLRH-HIVKRTLRLLQEREL 113
QY 703 VEPLTPSGAMPNOAKMILKETLRKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLR 762
DB 114 VEPLTPSGAMPNOAKMILKETLRKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKELR 173
QY 763 ENTSPKANKILDEAYVAVGVSFVYSRLIGICTSTVOLVQLMYPGCLLDHVENRQR 822
DB 174 ENTSPKANKILDEAYVAVGVSFVYSRLIGICTSTVOLVQLMYPGCLLDHVENRQR 233
QY 823 LGSQDLNMCQIAKGNLYLEDVRLVHRDLAANVLVKSPPNHVKITDFGLARLDIDETE 882
DB 234 LGSQDLNMCQIAKGNLYLEDVRLVHRDLAANVLVKSPPNHVKITDFGLARLDIDETE 293
QY 883 YHADGKVPYIKMALESILRRRTTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDL 942
DB 294 YHADGKVPYIKMALESILRRRTTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDL 353
QY 943 KGERLPORPCTIDVYIMVYKCMWIDSECRPRELVSFESRVARDPORFVYIQ-NEDLG 100
DB 354 KGERLPORPCTIDVYIMVYKCMWIDSECRPRELVSFESRVARDPORFVYIQ-NEDLG 413
QY 1002 PASPLDSTFYRSLLEDODMGDLVDAEYVLVPOGFCPPRPARAGAGMHHRRSSSTSG 106
DB 414 PASPLDSTFYRSLLEDODMGDLVDAEYVLVPOGFCPPRPARAGAGMHHRRSSSTSG 454
QY 1062 GGDITLGLERSEEARSP-----APSEGAGSDVFDGDLGMAKAGLSLPTHPSPLO 111
DB 455 -----SRTPLLSLSATSNNSATNCIDRNGG-----H----- 481
QY 1117 RYSEDPVPLPSETDGYVAFITCSPQPEYVQVPRQPPSPRGGPLPAARPAAGATLER 117
DB 482 -----PVREDEGL-----PAPRYVNO-LMPKKPSTANQVYINISLAIKSL 523
QY 1176 AKTISPKNGVAVDVFAPGAVENPEYL 1203
DB 524 LPIDSRYN-----SHSTAVDNPEYL 544

```

RESULT 14

S00727
 kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis
 C/Species: avian erythroblastosis virus
 C/Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
 C/Accession: S00727
 R/Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
 Oncogene Res. 1, 265-278, 1987
 A/Title: Common site of mutation in the erbB gene of avian erythroblastosis vi
 A/Reference number: S00727; M01D:88217326; P01D:2897102
 A/Accession: S00727

A:Molecule type: DNA
 A:Residues: 1-545 <SC0>
 A:Cross-references: EMBL:X06943
 C:Genetics:
 A:Gene: erbB
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; phosphotransferase
 F:135-400/Domain: protein kinase homology <KIN>
 F:143-151/Region: protein kinase ATP-binding motif

Query Match 23.9%; Score 1640; DB 2; Length 545;
 Best Local Similarity 54.9%; Pred. No. 14e-60;
 Matches 345; Conservative 69; Mismatches 122; Indels 92; Gaps 15;

QY 584 GREADQVACAHYKDPFCVACPSGVKPDLSYMPDKPDEGACPCPCINCHSCVLD 643
 DB 1 GP--DHQMKCAFIDQPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHPCTRGCKGKP 57
 QY 644 DDKGCPAEGRASPISIVAVY-GILVVVLGVVFGILIRKQOKIRKYMRLLOETEL 702
 DB 58 GLEGCP--NSKTSPIAGVVGSLCLVVGGLGLYLRR-HVKKRRLRLQEREL 113
 QY 703 VEPITPSGAMPNOAQMRILKETELRKVKVLSGAFGVYKGIWIPDGENYKIPVAIKYLR 762
 DB 114 VEPITPSGAPNOAHLRIKETEFKKVKVLGFGAGTVYKGLMPEGEKVTIIPVAIKELR 173
 QY 763 ENTSPKANEILDEAYVWAGSPYVSRLLGICLTSTVQLVTQMLPYGCLLDHYRENRGR 822
 DB 174 EATSPKANEILDEAYVWASVDNPHVCRLLGICLTSTVQLITQMLPYGCLLDYIREHKDN 233
 QY 823 LGSQDLNMCQIAGMSYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLRLDIDETE 882
 DB 234 IGSQYILNMCQIAGMNYLEERHLVHRDLAARNVLYKTPQDVKITDFGLAKOIGADEKE 293
 QY 883 YHADGKVPYIKMALESILRRFTHQSDVWSYGVTWELMTFGAKPYDGIIPAREIPDLLE 942
 DB 294 YHAEKGKVPYIKMALESILHRIYHQSDVWSYGVTWELMTFGSKPYDGIIPAREISVLE 353
 QY 943 KGERLPQRPICITIDYVMIMVKMWIDSECPREFRIVESESRMARDPQRFVILQ-NDLQ 1001
 DB 354 KGERLPQRPICITIDYVMIMVKMWSADSRPRFLIEFSKMRDPRLVLIQGDERRH 413
 QY 1002 PASPLDSTFYRSILDDDMGDLVDAEYLVQOGFFCPCDPAPAGAGVNHHRSSSTRSG 1061
 DB 414 IPSPTDSKFYRTLMEEDMEDIVDAEYLVPHQGF-----NSPT--- 454
 QY 1062 GGDILTLGLEPSEEARPSPL-----APSEGASDVFPDGLMGAKGLQSLPTHDPSEFLQ 1116
 DB 455 -----SRTPLLSSLSATSNNSATNCIDIRNG-----H----- 481
 QY 1117 RYSEDPYVLPSEITDGYVAPLTCSPQPEYVNOQDVRPQPPSPREGPLPAARPAQAT-LEK 1175
 DB 482 -----PYREDGFL-----PAPRYVNO--LMPKKPSTAMVNOQIYVISTLAISK 523
 QY 1176 AKTLSPGKGVKGVKVFARFAGAVENPEYL 1203
 DB 524 LPMDSRYON-----SHSTAVDNPEYL 544

RESULT 15

B44776
 protein-tyrosine kinase (BC 2.7.1.112) erbB - avian erythroblastosis virus (strain ES4)
 C:Species: avian erythroblastosis virus
 C:Date: 28-Apr-1993 #sequence_revision 28-Apr-1993 #text_change 04-Feb-2000
 C:Accession: B44776
 R:Bruskin, A.; Jackson, J.; Bishop, J.M.; McCauley, D.J.; Schatzman, R.C.
 Oncogene 5, 15-24, 1990
 A:Title: Six amino acids from the retroviral gene gag greatly enhance the transforming P
 A:Reference number: A44776; MUID:90206603; PMID:1969616
 A:Accession: B44776
 A:Status: Preliminary
 A:Molecule type: mRNA
 A:Residues: 1-540 <BRU>

A:Cross-references: GB:X52211
 C:Genetics:
 A:Gene: erbB
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; phosphotransferase; tyrosine-specific protein kinase
 F:130-395/Domain: protein kinase homology <KIN>
 F:138-146/Region: protein kinase ATP-binding motif

Query Match 23.7%; Score 1623; DB 2; Length 540;
 Best Local Similarity 54.9%; Pred. No. 7.1e-60;
 Matches 340; Conservative 69; Mismatches 120; Indels 90; Gaps 1

QY 593 CAHYDPPPCVACPSGVKPDLSYMPDKPDEGACPCPCINCHSCVLDLDDKCPAEQ 65
 DB 3 CAHFDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHPCTRGCKGPLEGCP--- 58
 QY 653 RASPLITSIVAVY-GILVVVLGVVFGILIRKQOKIRKYMRLLOETELVEFLTSGA 71
 DB 59 NSKTSPIAGVVGSLCLVVGGLGLYLRR-HVKKRRLRLQERELVEFLTSGE 11
 QY 712 MENOQMRILKETELRKVKVLSGAFGVYKGIWIPDGENYKIPVAIKVIRENTSPANK 77
 DB 118 AENOQHLRIKETEFKKVKVLGFGAGTVYKGLMPEGEKVTIIPVAIKELRENTSPANK 17
 QY 772 EILDAVYWAGSPYVSRLLGICLTSTVQLVTQMLPYGCLLDHYRENRGRSGQDLNM 83
 DB 178 EILDAVYWASVDNPHVCRLLGICLTSTVQLITQMLPYGCLLDYIREHKDNIGSQYILNM 23
 QY 832 CQIAGKMSYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLRLDIDETEYHADGKVP 89
 DB 238 CQIAGKMNYLEERHLVHRDLAARNVLYKTPQDVKITDFGLAKOIGADEKEYHAEKGKVP 29
 QY 892 IKMALESILRRFTHQSDVWSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQRP 95
 DB 298 IKMALESILHRIYHQSDVWSYGVTWELMTFGSKPYDGIIPAREISVLEKGERLPQRP 35
 QY 952 ICTIDVYVMIMVKMWIDSECPREFRIVESESRMARDPQRFVILQ-NDLQAPAPLSTF 10
 DB 358 ICTIDVYVMIMVKMWSADSRPRFLIEFSKMRDPRLVLIQGDERRHLPSTDSKF 41
 QY 1011 YRSILDDDMGDLVDAEYLVQOGFFCPCDPAPAGAGVNHHRSSSTRSGGDLTLGLE 10
 DB 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPT----- 44
 QY 1071 PSEEARPSPL-----APSEGASDVFPDGLMGAKGLQSLPTHDPSEFLQ 11
 DB 450 -----SRTPLLSSLSATSNNSATNCIDIRNG-----H----- 47
 QY 1126 LPSETDGYVAPLTCSPQPEYVNOQDVRPQPPSPREGPLPAARPAQAT-LEKAKTLSPGN 11
 DB 477 -PYREDGFL-----PAPRYVNO--LMPKKPSTAMVNOQIYVISTLAISKLPMDSRYON 52
 QY 1185 GVVKDVFARFAGAVENPEYL 1203
 DB 528 -----SHSTAVDNPEYL 539

Search completed: July 22, 2003, 09:05:47
 Job time : 31.0157 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2793 Seconds

(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-369-383-14

Perfect score: 6853

Sequence: 1 MEUAALCRGGLLALLPGA.....TFKGTPTANPEYGLDVPV 1261

Scoring table: ELOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6655	97.1	1355	1	ERB2_HUMAN
2	5859	85.5	1557	1	ERB2_RAT
3	5846.5	85.3	1254	1	ERB2_MESAU
4	3099	45.2	1210	1	EGFR_HUMAN
5	3076	44.9	1210	1	EGFR_MOUSE
6	2934.5	42.8	1308	1	ERB4_HUMAN
7	2915	42.5	1308	1	ERB4_RAT
8	2655.5	38.7	1167	1	XMRK_XIPMA
9	2391.5	34.9	1142	1	ERB3_HUMAN
10	2317.5	33.8	1339	1	ERB3_RAT
11	1935	28.2	1426	1	EGFR_DROME
12	1749.5	25.5	634	1	ERBB_ALV
13	1703	24.9	604	1	ERBB_AVIER
14	1630	23.8	540	1	ERBB_AVIEU
15	1558	22.7	703	1	EGFR_CHICK
16	1285	16.8	1323	1	L7R3_CAEEL
17	1142.5	16.7	245	1	ERB2_MOUSE
18	733	10.7	1363	1	ILPR_BRALA
19	696	10.2	1382	1	INSR_HUMAN
20	696	10.2	1607	1	MIPR_LYMGST
21	691	10.1	1383	1	INSR_RAT
22	690.5	10.0	1372	1	INSR_MOUSE
23	683	10.0	1300	1	IRRT_MOUSE
24	680	9.9	1477	1	HTK7_HYDAT
25	676	9.9	1297	1	IRRT_HUMAN
26	668.5	9.8	1300	1	IRRT_CAVPO
27	636	9.3	1367	1	IGIR_HUMAN
28	626	9.1	1390	1	INSR_AEDAE
29	625	9.1	1373	1	IGIR_MOUSE
30	621.5	9.0	1370	1	IGIR_RAT
31	615	8.0	2146	1	INSR_DROME
32	596.5	8.7	987	1	EPB4_HUMAN
33	588	8.6	1114	1	RET_HUMAN

34	585	8.5	976	1	EPB2_HUMAN	P29317 homo sapien
35	583.5	8.5	984	1	EPB1_CHICK	O07494 gallus gall
36	581.5	8.5	977	1	EPB2_MOUSE	Q00145 mus musculu
37	580.5	8.5	987	1	EPB4_MOUSE	P54761 mus musculu
38	577	8.4	757	1	HT16_HYDAT	P53356 hydra atten
39	574.5	8.4	984	1	EPB1_RAT	P09759 rattus norv
40	573.5	8.4	1053	1	PAK1_CHICK	Q00944 gallus gall
41	573	8.4	902	1	EPB8_XENLA	Q91736 xenopus lae
42	569	8.3	1068	1	PAK1_XENLA	Q91738 xenopus lae
43	568.5	8.3	984	1	EPB1_HUMAN	P54762 homo sapien
44	567.5	8.3	985	1	EPB4_XENLA	O91571 xenopus lae
45	563	8.2	1052	1	PAK1_MOUSE	P34152 mus musculu

ALIGNMENTS

RESULT 1
ERB2_HUMAN STANDARD; PRT; 1255 AA.
ID ERB2_HUMAN
AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE surface receptor HER2) (MLN 19).
GN ERBB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86118663; PubMed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA Saito T., Toyoshima K.,
RT "Similarity of protein encoded by the human c-erbB-2 gene to
RT epidermal growth factor receptor.";
RL Nature 319:230-234(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86070181; PubMed=2999974;
RA Cousens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA McGrath U., Seeburg P.H., Libermann T.A., Schlessinger J.,
RA Franke U., Levinson A., Ulrich A.,
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene.";
RL Science 230:1132-1139(1985).
RN [3]
RP SEQUENCE OF 727-1031 FROM N.A.
RX MEDLINE=86016729; PubMed=2995967;
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.,
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor receptor gene and is amplified in a
RT human salivary gland adenocarcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN [4]
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=809548;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.,
RT "Characterization of a new allele of the human ERBB2 gene by allele-
RT specific competition hybridization.";
RL Genomics 15:426-429(1993).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMBIREGULIN.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN
CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL curation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M11767; AAA35808.1; -.
DR EMBL: M11761; AAA35808.1; JOINED.
DR EMBL: M11762; AAA35808.1; JOINED.
DR EMBL: M11763; AAA35808.1; JOINED.
DR EMBL: M11764; AAA35808.1; JOINED.
DR EMBL: M11765; AAA35808.1; JOINED.
DR EMBL: M11766; AAA35808.1; JOINED.
DR EMBL: M11730; AAA75493.1; -.
DR EMBL: M12036; AAA35828.1; -.
DR EMBL: X03363; CA427060.1; -.
DR PIR: A25491; A25491.
DR PIR: A24571; A24571.
DR HSSP: P11362; IFGK.
DR Genew: HGNC:3430; ERBB2.
DR MIM: 164870; -.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_Pkinase.
DR InterPro: IPR004019; YLP motif.
DR Pfam: PF00069; Pkinase: 1.
DR Pfam: PF00757; Furin-like: 1.
DR Pfam: PF01030; Recep_L_domain: 2.
DR Pfam: PF02757; YLP: 2.
DR Pfam: PD000001; Euk_Pkinase: 1.
DR SMART: SM00261; FU: 3.
DR SMART: SM00219; Tyrc: 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP: 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR: 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM: 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transmembrane; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Polymorphism.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 653 675 POTENTIAL.
FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 720 987 PROTEIN KINASE.
FT NP_BIND 726 734 ATP (BY SIMILARITY).
FT BINDING 753 753 ATP (BY SIMILARITY).
FT ACT_SITE 845 845 BY SIMILARITY.
FT DISULFID 195 204 BY SIMILARITY.
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 220 227 BY SIMILARITY.
FT DISULFID 224 227 BY SIMILARITY.
FT DISULFID 224 235 BY SIMILARITY.
FT DISULFID 236 244 BY SIMILARITY.
FT DISULFID 240 252 BY SIMILARITY.
FT DISULFID 255 264 BY SIMILARITY.
FT DISULFID 268 295 BY SIMILARITY.
FT DISULFID 299 311 BY SIMILARITY.
FT DISULFID 315 331 BY SIMILARITY.
FT DISULFID 334 338 BY SIMILARITY.
FT DISULFID 511 520 BY SIMILARITY.
FT DISULFID 515 528 BY SIMILARITY.
FT DISULFID 531 540 BY SIMILARITY.
FT DISULFID 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.
FT DISULFID 567 584 BY SIMILARITY.
FT DISULFID 587 596 BY SIMILARITY.
FT DISULFID 600 623 BY SIMILARITY.
FT DISULFID 626 634 BY SIMILARITY.
FT DISULFID 630 642 BY SIMILARITY.
FT MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 124 124 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 259 259 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 530 530 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 571 571 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).
FT VARIANT 654 654 I -> V.
FT VARIANT 654 654 I -> V.
FT VARIANT 655 655 I -> V.
FT VARIANT 655 655 I -> V.
FT CONFLICT 1170 1170 /FTID=VAR_004078.
FT SEQUENCE 1255 AA; 137909 MM; 39E9DEDA04DC6962 CRC64;
SQ
Query Match 97.1%; Score 6655; DB 1; Length 1255;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1229; Conservative 11; Mismatches 14; Indels 8; Gaps 2
1 MELALCRMGILLALPPGAASSTQVCTGDMKRLPASPETHLDMRLHYGGGVVQGNL 60
1 MELALCRMGILLALPPGAASSTQVCTGDMKRLPASPETHLDMRLHYGGGVVQGNL 60
61 ELTYLPTNASLSFLQDIOEVQSVLLIAHQVROVPLQRLIRVGTQLFEDNYALAVLNG 120
61 ELTYLPTNASLSFLQDIOEVQSVLLIAHQVROVPLQRLIRVGTQLFEDNYALAVLNG 120
61 ELTYLPTNASLSFLQDIOEVQSVLLIAHQVROVPLQRLIRVGTQLFEDNYALAVLNG 120
121 DPANNTPTVAGSPGGRRLQSLTEILKGGTLQRLNQLCYQDTILMKDIFHKNNQLA 180
121 DPANNTPTVAGSPGGRRLQSLTEILKGGTLQRLNQLCYQDTILMKDIFHKNNQLA 180
121 DPANNTPTVAGSPGGRRLQSLTEILKGGTLQRLNQLCYQDTILMKDIFHKNNQLA 180
121 DPANNTPTVAGSPGGRRLQSLTEILKGGTLQRLNQLCYQDTILMKDIFHKNNQLA 180
181 LTLIDITNSRACHPCSPMCKGRSMWESSHDCSLTRTYAGGACARCKPLPTDCHEQC 240
181 LTLIDITNSRACHPCSPMCKGRSMWESSHDCSLTRTYAGGACARCKPLPTDCHEQC 240
181 LTLIDITNSRACHPCSPMCKGRSMWESSHDCSLTRTYAGGACARCKPLPTDCHEQC 240
181 LTLIDITNSRACHPCSPMCKGRSMWESSHDCSLTRTYAGGACARCKPLPTDCHEQC 240
241 AAGGTGKSHDCLACHFNHSGICEHLCPALVYNTDTFESMPNREGRYTFGASCTYACP 300
241 AAGGTGKSHDCLACHFNHSGICEHLCPALVYNTDTFESMPNREGRYTFGASCTYACP 300
241 AAGGTGKSHDCLACHFNHSGICEHLCPALVYNTDTFESMPNREGRYTFGASCTYACP 300
241 AAGGTGKSHDCLACHFNHSGICEHLCPALVYNTDTFESMPNREGRYTFGASCTYACP 300
301 YNYLSTVGSCTIVCPLEHNOEVTAEADGTORCEKSCPCARVCGYGLMFPNFTVFWLR-V 359
301 YNYLSTVGSCTIVCPLEHNOEVTAEADGTORCEKSCPCARVCGYGLMFPNFTVFWLR-V 359
301 YNYLSTVGSCTIVCPLEHNOEVTAEADGTORCEKSCPCARVCGYGLMFPNFTVFWLR-V 359
301 YNYLSTVGSCTIVCPLEHNOEVTAEADGTORCEKSCPCARVCGYGLMFPNFTVFWLR-V 359
360 PKYSASHLEFPAGCKKIFGSLAFLPSPFGDDPASNTAPQYIKANSKFITELTGYLYIS 419
360 PKYSASHLEFPAGCKKIFGSLAFLPSPFGDDPASNTAPQYIKANSKFITELTGYLYIS 419
360 PKYSASHLEFPAGCKKIFGSLAFLPSPFGDDPASNTAPQYIKANSKFITELTGYLYIS 419
360 PKYSASHLEFPAGCKKIFGSLAFLPSPFGDDPASNTAPQYIKANSKFITELTGYLYIS 419
354 RAYTSANIOEFACKCKIFGSLAFLPSPFGDDPASNTAPQYIKANSKFITELTGYLYIS 413
354 RAYTSANIOEFACKCKIFGSLAFLPSPFGDDPASNTAPQYIKANSKFITELTGYLYIS 413
354 RAYTSANIOEFACKCKIFGSLAFLPSPFGDDPASNTAPQYIKANSKFITELTGYLYIS 413
354 RAYTSANIOEFACKCKIFGSLAFLPSPFGDDPASNTAPQYIKANSKFITELTGYLYIS 413
420 AMPPSLDLSPVONLQVITGRILHNGAVSLTLOGIGISLGLRSIRREGSGALIHNNH 479
420 AMPPSLDLSPVONLQVITGRILHNGAVSLTLOGIGISLGLRSIRREGSGALIHNNH 479
420 AMPPSLDLSPVONLQVITGRILHNGAVSLTLOGIGISLGLRSIRREGSGALIHNNH 479
420 AMPPSLDLSPVONLQVITGRILHNGAVSLTLOGIGISLGLRSIRREGSGALIHNNH 479
414 AMPPSLDLSPVONLQVITGRILHNGAVSLTLOGIGISLGLRSIRREGSGALIHNNH 473
414 AMPPSLDLSPVONLQVITGRILHNGAVSLTLOGIGISLGLRSIRREGSGALIHNNH 473
414 AMPPSLDLSPVONLQVITGRILHNGAVSLTLOGIGISLGLRSIRREGSGALIHNNH 473
414 AMPPSLDLSPVONLQVITGRILHNGAVSLTLOGIGISLGLRSIRREGSGALIHNNH 473
480 LCPVHTVPMQDLFRNPHQALHTANRPEDECVGEGLAGHQLCARGHGCMGPGPTQCVNSQ 539
480 LCPVHTVPMQDLFRNPHQALHTANRPEDECVGEGLAGHQLCARGHGCMGPGPTQCVNSQ 539
480 LCPVHTVPMQDLFRNPHQALHTANRPEDECVGEGLAGHQLCARGHGCMGPGPTQCVNSQ 539
480 LCPVHTVPMQDLFRNPHQALHTANRPEDECVGEGLAGHQLCARGHGCMGPGPTQCVNSQ 539
474 LCPVHTVPMQDLFRNPHQALHTANRPEDECVGEGLAGHQLCARGHGCMGPGPTQCVNSQ 533
474 LCPVHTVPMQDLFRNPHQALHTANRPEDECVGEGLAGHQLCARGHGCMGPGPTQCVNSQ 533
474 LCPVHTVPMQDLFRNPHQALHTANRPEDECVGEGLAGHQLCARGHGCMGPGPTQCVNSQ 533
474 LCPVHTVPMQDLFRNPHQALHTANRPEDECVGEGLAGHQLCARGHGCMGPGPTQCVNSQ 533
540 FLNGQEVVEECRYLQGLPREYVNAARCLSCHEQCQPNQNSVTCFPGPADQVACAHYNDP 599
540 FLNGQEVVEECRYLQGLPREYVNAARCLSCHEQCQPNQNSVTCFPGPADQVACAHYNDP 599
540 FLNGQEVVEECRYLQGLPREYVNAARCLSCHEQCQPNQNSVTCFPGPADQVACAHYNDP 599
540 FLNGQEVVEECRYLQGLPREYVNAARCLSCHEQCQPNQNSVTCFPGPADQVACAHYNDP 599
534 FLNGQEVVEECRYLQGLPREYVNAARCLSCHEQCQPNQNSVTCFPGPADQVACAHYNDP 593
534 FLNGQEVVEECRYLQGLPREYVNAARCLSCHEQCQPNQNSVTCFPGPADQVACAHYNDP 593
534 FLNGQEVVEECRYLQGLPREYVNAARCLSCHEQCQPNQNSVTCFPGPADQVACAHYNDP 593
534 FLNGQEVVEECRYLQGLPREYVNAARCLSCHEQCQPNQNSVTCFPGPADQVACAHYNDP 593
600 PFCVACPSGVKXKDLSTYMPIMKPEDEGACQPCPINCTHSQVLDLDDKGPAPORASPLTS 659
600 PFCVACPSGVKXKDLSTYMPIMKPEDEGACQPCPINCTHSQVLDLDDKGPAPORASPLTS 659
600 PFCVACPSGVKXKDLSTYMPIMKPEDEGACQPCPINCTHSQVLDLDDKGPAPORASPLTS 659
600 PFCVACPSGVKXKDLSTYMPIMKPEDEGACQPCPINCTHSQVLDLDDKGPAPORASPLTS 659
594 PFCVACPSGVKXKDLSTYMPIMKPEDEGACQPCPINCTHSQVLDLDDKGPAPORASPLTS 653
594 PFCVACPSGVKXKDLSTYMPIMKPEDEGACQPCPINCTHSQVLDLDDKGPAPORASPLTS 653
594 PFCVACPSGVKXKDLSTYMPIMKPEDEGACQPCPINCTHSQVLDLDDKGPAPORASPLTS 653
594 PFCVACPSGVKXKDLSTYMPIMKPEDEGACQPCPINCTHSQVLDLDDKGPAPORASPLTS 653
660 IYAVVIGILLVVLGVVFGILLKRRQKIRKTYMRLLOETELVEPLTPSGAMPNOQMR 719
660 IYAVVIGILLVVLGVVFGILLKRRQKIRKTYMRLLOETELVEPLTPSGAMPNOQMR 719
660 IYAVVIGILLVVLGVVFGILLKRRQKIRKTYMRLLOETELVEPLTPSGAMPNOQMR 719
660 IYAVVIGILLVVLGVVFGILLKRRQKIRKTYMRLLOETELVEPLTPSGAMPNOQMR 719
654 IYAVVIGILLVVLGVVFGILLKRRQKIRKTYMRLLOETELVEPLTPSGAMPNOQMR 713
654 IYAVVIGILLVVLGVVFGILLKRRQKIRKTYMRLLOETELVEPLTPSGAMPNOQMR 713
654 IYAVVIGILLVVLGVVFGILLKRRQKIRKTYMRLLOETELVEPLTPSGAMPNOQMR 713
654 IYAVVIGILLVVLGVVFGILLKRRQKIRKTYMRLLOETELVEPLTPSGAMPNOQMR 713

```

OY 720 ILKETEARKVVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLENTSPXANKIIDEAVY 779
DB 714 ILKETEARKVVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLENTSPXANKIIDEAVY 773
OY 780 MAGVGSPIVSRLLGICITSTVQVLTQMPYGCILLDHYREKRLGSGODLLNMCQIAKAM 839
DB 774 MAGVGSPIVSRLLGICITSTVQVLTQMPYGCILLDHYREKRLGSGODLLNMCQIAKAM 833
OY 840 SYLEDVRLVHRDLAARVVLVKSPIHVKITDFFGLARLLDIDETFYHADGGKVPKMALES 899
DB 834 SYLEDVRLVHRDLAARVVLVKSPIHVKITDFFGLARLLDIDETFYHADGGKVPKMALES 893
OY 900 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPPICITIDVYM 959
DB 894 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPPICITIDVYM 953
OY 960 IMVKMMIDSECRPRFELVSEFSRMARDPQRFVVIQNEIDLGASPLDSTFYRSLLDDDD 1019
DB 954 IMVKMMIDSECRPRFELVSEFSRMARDPQRFVVIQNEIDLGASPLDSTFYRSLLDDDD 1013
OY 1020 MGDVDAEEYLVPOQGFPCDPAPAGAGVHHRRSSSTRSGGDLTLGLEPSEEPAPRS 1079
DB 1014 MGDVDAEEYLVPOQGFPCDPAPAGAGVHHRRSSSTRSGGDLTLGLEPSEEPAPRS 1073
OY 1080 PLASEGAGSDVPFGDILGMAAKGLQSLPTHDSPLQRYSEDPVPLPSETDGYAPLTC 1139
DB 1074 PLASEGAGSDVPFGDILGMAAKGLQSLPTHDSPLQRYSEDPVPLPSETDGYAPLTC 1133
OY 1140 SPOPEYVNCPPVRPQPPSPREGPLPAARPAATLEBAKTISPKNQVNVQDVAFGAGAVEN 1199
DB 1134 SPOPEYVNCPPVRPQPPSPREGPLPAARPAATLEBAKTISPKNQVNVQDVAFGAGAVEN 1193
OY 1200 PEVYLTPOGGAAPQHPAPPAPSPAFDNLYVWDQPPERGAPOSTFKCTPTAENPEVYGLDV 1259
DB 1194 PEVYLTPOGGAAPQHPAPPAPSPAFDNLYVWDQPPERGAPOSTFKCTPTAENPEVYGLDV 1253
OY 1260 PV 1261
DB 1254 PV 1255

RESULT 2
ID ERB2_RAT STANDARD; PRT; 1257 AA.
AC P06454;
DT 01-JAN-1988 (Rel. 06, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbb-2 precursor (BC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor
DE receptor-related protein).
CN ERBB2 OR NEU.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroblastoma;
RX MEDLINE=66118662; PubMed=3945311;
RA Bargman C.I., Hung M.-C., Weinberg R.A.;
RT "The neu oncogene encodes an epidermal growth factor receptor-related
RT protein.";
RL Nature 319:226-230(1986).
RN 12
RP SEQUENCE OF 852-905 FROM N.A.
RC TISSUE=Sclerotic nerve;
RX MEDLINE=9122560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).

```

```

RN 895
RP STRUCTURE BY NMR OF 650-668.
RX MEDLINE=92155181; PubMed=1346763;
RA Gullick W.J., Bottomley A.C., Lofte F.J., Doak D.G., Mulvey D.,
RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
RT "Three dimensional structure of the transmembrane region of the proto-
RT oncogenic and oncogenic forms of the neu protein.";
RL EMO J. 11:43-48(1992).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboratio
CC between the Swiss Institute of Bioinformatics and the EMBL databank
CC The European Bioinformatics Institute. There are no restrictions on it
CC use by non-profit institutions as long as its content is in no wa
CC modified and this statement is not removed. Usage by and for commercia
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X03362; CAA27059.1; AUL_INIT.
DR PIR: A24562; TYRNTU.
DR HSSP: P11362; IPRGX.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_kinase.
DR InterPro: IPR004019; YLP_motif.
DR Pfam: PF00066; Kinase; 1.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR ProDom: PD000001; Euk_kinase; 1.
DR SMART: SM00261; Fu_3.
DR SMART: SM00219; Tyrc; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
DR Transferase; Tyrosine-protein Kinase; ATP-binding; Phosphorylation;
KW Proto-oncogene; Disease mutation.
FT SIGNAL 1..21
FT CHAIN 22..1257
FT DOMAIN 22..654
FT TRANSMEM 655..677
FT DOMAIN 678..1257
FT DOMAIN 159..369
FT DOMAIN 473..646
FT DOMAIN 722..989
FT NP_BIND 728..736
FT BINDING 755..755
FT ACT_SITE 847..847
FT DISULFID 196..205
FT DISULFID 200..213
FT DISULFID 221..228
FT DISULFID 225..236
FT DISULFID 237..245
FT DISULFID 241..253
FT DISULFID 256..265
FT DISULFID 269..296
FT DISULFID 300..312
FT DISULFID 316..332
FT DISULFID 335..339
FT DISULFID 513..522

```

FT DISULFID 517 530 BY SIMILARITY.
 FT DISULFID 533 542 BY SIMILARITY.
 FT DISULFID 546 562 BY SIMILARITY.
 FT DISULFID 565 578 BY SIMILARITY.
 FT DISULFID 589 586 BY SIMILARITY.
 FT DISULFID 598 598 BY SIMILARITY.
 FT DISULFID 602 625 BY SIMILARITY.
 FT DISULFID 628 636 BY SIMILARITY.
 FT MOD. RES 1141 1141 BY SIMILARITY.
 FT MOD. RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 188 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
 SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 85.5%; Score 5859; DB 1; Length 1257;
 Best Local Similarity 85.8%; Pred. No. 6,5e-306;
 Matches 1084; Conservative 58; Mismatches 112; Indels 10; Gaps 4;

QY 1 METALCRMGLLALLPFGASTQVCTGTDKMLRPLASPEHLDMLRLHYOGQVYQGNL 60
 DB 1 METALCRMGLLALLPFGIAGTQVCTGTDKMLRPLASPEHLDMLRLHYOGQVYQGNL 60
 QY 61 ELTYLPTNASLFLDIOEVQYVLIANQVRYQLRLRIVREGTQLFEDNVALAVLDNG 120
 DB 61 ELTYVPAVASLFLDIOEVQYVLIANQVRYQLRLRIVREGTQLFEDNVALAVLDNR 120
 QY 121 DPLANTTPTV-GASGGIREIQLSLTETILKGVLIQNPOLCYQDTILKDIIPHKNOL 179
 DB 121 DPQNVASTPGRTEGRLRELQLSLTEILKGVLIQNPOLCYQDMVLMKDVFRKNOL 180
 QY 180 ALLIDNRSRACHPSCPMCKSRCKSSSEDCSLRTRVAGGACACKGPLPTDCHEQ 239
 DB 181 APVIDIDNRSRACHPSCACACKDNCKGSPEDCQILITICTSCACACKRLPTDCHEQ 240
 QY 240 CAAGCTGPRHSDCLACLFHNSGICELCPALVTYNTDTESMNPGRYTFGASCVTAC 299
 DB 241 CAAGCTGPRHSDCLACLFHNSGICELCPALVTYNTDTESMNPGRYTFGASCVTTC 300
 QY 300 PYNVLSTDVGSCTLVCPILHNOEVAEDGTORCEKCSKPCARVCYGLGMPNFTVSPRLR 359
 DB 301 PYNVLSTEVGCTLVCPNNOEVAEDGTORCEKCSKPCARVCYGLGMEH-----LKG 353
 QY 360 PK-VASHLSEEPAGCKIFGSLAFLPESFDGDPASNTAPQYIKANSKFIGITELTGLYI 418
 DB 364 ARAITSDNVQFEDGCKIFGSLAFLPESFDGDPSSG:APLRPEQLQVFETLEITGLYI 413
 QY 419 SAWPDLSDLVFQNLQYIRRIILHNKAYSLTLOGISLWGLRLSELGSLALIHNT 478
 DB 414 SAWPDLSDLVFQNLRIIRRIILHDAVSLTLOGISLWGLRLSELGSLALIHNT 473
 QY 479 HLCFVHTVPMQDLFNNPQALHNTANRPEDE-CVGEGLACHOLCARHCGWGPPTQVNC 537
 DB 474 HLCFVHTVPMQDLFNNPQALHNGNREBBDLCVSSSLVNSLCAHCHCGPPTQVNC 533
 QY 538 SQFLRGQCEVECRVLQGLPREYVNAHCLPCHCECPQNGSVTCGPEPADQCAAHYK 597
 DB 534 SHFLRGQCEVECRVWKLPREYVSDKRLPCHCECPQNGSVTCGPEPADQCAAHYK 593
 QY 598 DPFCVACSPGVAPDLSYPIWKFPDEEGACQCPINTCHSCYVDLDDKCPAQRASPL 657
 DB 594 DSSCVACSPGVAPDLSYPIWKFPDEEGACQCPINTCHSCYVDLDDKCPAQRASPL 653
 QY 658 TSISAVVGIILVVVLGVFGLIKRQOKRKYTMRLIQETELVEPLTPSGAMPNQAQ 717
 DB 654 TSISAVVGIILVVVLGVFGLIKRQOKRKYTMRLIQETELVEPLTPSGAMPNQAQ 713
 QY 718 MRLKETELRKVKVLGSAFGTVYKGIWIPDGEVVKIPVAIXVLRNTSPKANKELIDEA 777

DB 714 MRLKETELRKVKVLGSAFGTVYKGIWIPDGEVVKIPVAIXVLRNTSPKANKELIDEA 773
 QY 778 YVMAGVSPYRLLIGLITSTVOLVQMLPBYGLDHYENRGRGLSQDLNMCQIAK 833
 DB 774 YVMAGVSPYRLLIGLITSTVOLVQMLPBYGLDHYENRGRGLSQDLNMCQIAK 833
 QY 838 GMSYEDVRLVHRLDAAARNVYKPNVVKITDPGLARLLIDETRYADGKYPKMMAL 897
 DB 834 GMSYEDVRLVHRLDAAARNVYKPNVVKITDPGLARLLIDETRYADGKYPKMMAL 893
 QY 898 ESILRRRTTHSGDVSYGVTYVWELMTFGAARYDGIIPARZIPDLLEKGERLPQPPICITDV 957
 DB 894 ESILRRRTTHSGDVSYGVTYVWELMTFGAARYDGIIPARZIPDLLEKGERLPQPPICITDV 953
 QY 958 YMIWVKCMIDSECRPPRELVEFSFMAADPQRFVYIQNEDIQPSPLUSTYRSLLED 101
 DB 954 YMIWVKCMIDSECRPPRELVEFSFMAADPQRFVYIQNEDIQPSPLUSTYRSLLED 101
 QY 1018 DDMGDLVDAEYLVPOQFPCCPDAPAGGVVHRRSSSTRSGQDPLTGLEPSEBEAP 107
 DB 1014 DDMGDLVDAEYLVPOQFPSPDPTPGTARHRRSSSTRSGQDPLTGLEPSEBEAP 107
 QY 1078 RSLPAPSGASDVFDGDLGKAAGKIOSLPTHDPSPLQYSEDPTVPLFSETDGYAPL 113
 DB 1074 RSLPAPSGASDVFDGDLGKAAGKIOSLPTHDPSPLQYSEDPTVPLFSETDGYAPL 113
 QY 1138 TCSPOPEYVNPQDVPPOPSPREGPLPAABPAGATLEBATLSFGKNGVYKDYAFAGAV 115
 DB 1134 ACSPOPEYVNPQDVPPOPSPREGPLPAABPAGATLEBATLSFGKNGVYKDYAFAGAV 115
 QY 1194 ENPEYLVFREGTASPPHSPAFSPFNLYWQNSSEQPPSPNFEGETPAENPEYGL 125
 DB 1194 ENPEYLVFREGTASPPHSPAFSPFNLYWQNSSEQPPSPNFEGETPAENPEYGL 125
 QY 1258 DVPV 1261
 DB 1254 DVPV 1257

RESULT 3
 ERB2_MESAU STANDARD; PRT; 1254 AA.
 AC 060533;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
 DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
 GN ERB2 OR NEU.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Neve;
 RX MEDLINE=94153007; PubMed=7908275;
 RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
 RA Yamazaki Y., Ishikawa T.;
 RT "Cloning and activation of the Syrian hamster neu proto-oncogene."
 RL Gene 140:251-255(1994).
 CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
 ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
 POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
 ALPHA AND AMPHIREGULIN (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 tyrosine phosphate.
 CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 (POTENTIAL).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE

CC RESIDUES.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D16295; BAA03801.1; -
 DR HSPB; P11362; IREK.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_Pkinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00069; Pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR Pfam: PF02757; YLP; 2.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SMC0261; FU; 3.
 DR SMART: SMC0219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Proto-oncogene; Disease mutation.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 1254 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
 FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 653 675 POTENTIAL.
 FT DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 158 368 CYS-RICH.
 FT DOMAIN 472 644 CYS-RICH.
 FT NP_BIND 720 987 PROTEIN KINASE.
 FT BINDING 726 734 ATP (BY SIMILARITY).
 FT ACT_SITE 753 753 ATP (BY SIMILARITY).
 FT ACT_SITE 845 845 BY SIMILARITY.
 FT DISULFID 195 204 BY SIMILARITY.
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 236 244 BY SIMILARITY.
 FT DISULFID 240 252 BY SIMILARITY.
 FT DISULFID 255 264 BY SIMILARITY.
 FT DISULFID 268 295 BY SIMILARITY.
 FT DISULFID 299 311 BY SIMILARITY.
 FT DISULFID 315 331 BY SIMILARITY.
 FT DISULFID 334 338 BY SIMILARITY.
 FT DISULFID 511 520 BY SIMILARITY.
 FT DISULFID 515 528 BY SIMILARITY.
 FT DISULFID 540 540 BY SIMILARITY.
 FT DISULFID 544 560 BY SIMILARITY.
 FT DISULFID 563 576 BY SIMILARITY.
 FT DISULFID 584 584 BY SIMILARITY.
 FT DISULFID 587 596 BY SIMILARITY.
 FT DISULFID 600 623 BY SIMILARITY.
 FT DISULFID 626 634 BY SIMILARITY.
 FT DISULFID 630 642 BY SIMILARITY.
 FT MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 125 125 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 187 187 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 259 259 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 530 530 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 571 571 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 658 658 V -> E (IN ONCOGENIC NEU).
 FT VARIANT 659 659 V -> E (IN ONCOGENIC NEU).
 SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

Query Match 85.3%; Score 5846.5; DB 1; Length 1254;
 Best Local Similarity 85.3%; Pred. No. 3e-105;
 Matches 1077; Conservative 69; Mismatches 107; Indels 9; Gaps 3

QY	1	MELALCRWGLLALPPGAASQVCTGDMKRLPASPETHLMDRLHYOGCQVQSNL	60
DB	1	MELAAWGWGLLALLPAGASQVCTGDMKRLPASPETHLMDRLHYOGCQVQSNL	60
QY	61	ELTLPNNAISFLQDIQEVGVVLLAHNOVRQVPLQRIATVGTGLFEDNYLAVLDNG	120
DB	61	ELTLPNNAISFLQDIQEVGVVLLAHNOVRQVPLQRIATVGTGLFEDNYLAVLDNR	120
QY	121	DLNNNTPVVGASPGRLRELQRLSLTEILKGVYLQRPOLCYQDTILMKDIFPKKNOLA	180
DB	121	DLNNNTPVVGASPGRLRELQRLSLTEILKGVYLQRPOLCYQDTILMKDIFPKKNOLA	180
QY	181	LTLDTNRSRACHGSCMCKGRCKGSSSDQSLRTVAGGACACKGPLPTDCCHEOC	240
DB	181	LTLDTNRSRACHGSCMCKGRCKGSSSDQSLRTVAGGACACKGPLPTDCCHEOC	240
QY	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPFGRTYFGASCYTACP	300
DB	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPFGRTYFGASCYTACP	300
QY	301	YNYLSTVGSCTIVCPILHNOEVTAEADGTQCEKCSKPCARVCYGLMFPNFVSWLRVP	360
DB	301	YNYLSTVGSCTIVCPILHNOEVTAEADGTQCEKCSKPCARVCYGLMFPNFVSWLRVP	360
QY	361	K-VASHLKEPAGCKKIFGSLATLPESFPDDPSNPAQYIKANSFGICTELTYGLYS	419
DB	361	K-VASHLKEPAGCKKIFGSLATLPESFPDDPSNPAQYIKANSFGICTELTYGLYS	419
QY	420	AMPDLPDLVSFQNLQVIRGRILHNGAYSLTLOGLSIMLGRSLRELSGALLIHNTH	479
DB	420	AMPDLPDLVSFQNLQVIRGRILHNGAYSLTLOGLSIMLGRSLRELSGALLIHNTH	479
QY	479	LCFHTHPMQLPNNPQALLHTANRPEDECVBEGALCHQLCARHGCKGPPQVCNCSG	539
DB	479	LCFHTHPMQLPNNPQALLHTANRPEDECVBEGALCHQLCARHGCKGPPQVCNCSG	539
QY	539	FLRQGECEVEGRVLOGLPREYVABHCLPCHPECQONGSVTCFGEADQCVACAYKDP	599
DB	539	FLRQGECEVEGRVLOGLPREYVABHCLPCHPECQONGSVTCFGEADQCVACAYKDP	599
QY	599	FLRQGECEVEGRVLOGLPREYVABHCLPCHPECQONGSVTCFGEADQCVACAYKDP	599
DB	599	FLRQGECEVEGRVLOGLPREYVABHCLPCHPECQONGSVTCFGEADQCVACAYKDP	599
QY	659	PFVAPRCPGKEDLSYMPIMKYPDEEGMCPICNTHSCVDLDERGCPAEGRASPLTS	659
DB	659	PFVAPRCPGKEDLSYMPIMKYPDEEGMCPICNTHSCVDLDERGCPAEGRASPLTS	659
QY	719	IVSAVVGILLVVLGVVFGILLIRROCKIRKYMRLLOETELVEBLTSGAMPNOQNR	719
DB	719	IVSAVVGILLVVLGVVFGILLIRROCKIRKYMRLLOETELVEBLTSGAMPNOQNR	719
QY	779	ILKETELRKVKVLSGAGFTVYKGIWIPQENYKIPVAIKVIRENTSPRANKIILDEAV	779
DB	779	ILKETELRKVKVLSGAGFTVYKGIWIPQENYKIPVAIKVIRENTSPRANKIILDEAV	779
QY	839	MAGGSGYVYRLLGICLSTVQVLTQLMRPGCLLDHVRNKGRLSGQDILNNCQJAKKM	839
DB	839	MAGGSGYVYRLLGICLSTVQVLTQLMRPGCLLDHVRNKGRLSGQDILNNCQJAKKM	839
QY	899	SYLEDVAVLVRDLAARVLYKSPNHKITDFGLARLLDIDETEHADGGKVPKIMNALES	899
DB	899	SYLEDVAVLVRDLAARVLYKSPNHKITDFGLARLLDIDETEHADGGKVPKIMNALES	899
QY	959	ILRRFTHQSDVWSYGTWELMTFAKPYDGIIPAREIDPLEKGRLPQPICTIDVYM	959
DB	959	ILRRFTHQSDVWSYGTWELMTFAKPYDGIIPAREIDPLEKGRLPQPICTIDVYM	959
QY	1019	IMYKMWIDEECPRELYSESRVARDPQFVYVQNEIDLPAASLDSTFYSRLDEDD	1019
DB	1019	IMYKMWIDEECPRELYSESRVARDPQFVYVQNEIDLPAASLDSTFYSRLDEDD	1019
QY	1077	IMYKMWIDEECPRELYSESRVARDPQFVYVQNEIDLPAASLDSTFYSRLDEDD	1077
DB	1077	IMYKMWIDEECPRELYSESRVARDPQFVYVQNEIDLPAASLDSTFYSRLDEDD	1077

RT and supercoiled DNA.";
 RN Nature 309:270-273(1984).
 RN (15)
 RP PHOSPHORYLATION.
 RX MEDLINE=89278137; PubMed=2543678;
 RA Margolis B.L., Lax I., Kris R., Dombalagian M., Hongger A.M.,
 RA Hawk R., Givol D., Ullrich A., Schlessinger J.;
 RT "All autophosphorylation sites of epidermal growth factor (EGF)
 RT receptor and HER2/neu are located in their carboxyl-terminal tails.
 RT Identification of a novel site in EGF receptor.";
 RL J. Biol. Chem. 264:10667-10671(1989).
 RN [16]
 RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
 RP ASN-528.
 RX MEDLINE=96398132; PubMed=8962717;
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
 RT "Analysis of the glycosylation patterns of the extracellular domain of
 RT the epidermal growth factor receptor expressed in Chinese hamster
 RT ovary fibroblasts.";
 RL Growth Factors 13:121-132(1996).
 RN [17]
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
 RP ASN-603.
 RX MEDLINE=20198209; PubMed=10731689;
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
 RT "Characterization of the N-oligosaccharides attached to the atypical
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor
 RT receptor.";
 RL J. Biochem. 127:65-72(2000).
 RN [18]
 RP PARTIAL SEQUENCE AND DISULFIDE BONDS.
 RX MEDLINE=98225196; PubMed=9556602;
 RA Abe Y., Otake M., Inasaki F., Lax I., Schlessinger J., Kohda D.;
 RT "Disulfide bond structure of human epidermal growth factor receptor.";
 RL J. Biol. Chem. 273:11150-11157(1998).
 RN [19]
 RP REVIEW.
 RX MEDLINE=87297456; PubMed=3039909;
 RA Carpenter G.;
 RT "Receptors for epidermal growth factor and other polypeptide
 RT mitogens.";
 RL Annu. Rev. Biochem. 56:881-914(1987).
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
 CC EGF-like growth factor, gp30 and vaccinia virus growth factor. Is
 CC involved in the control of cell growth and differentiation.
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
 CC secreted.
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/
 CC truncated isoform/TEGFR, 3/p110 and 4, are produced by
 CC alternative splicing.
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
 CC expressed in ovarian cancers.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; X00588; CA25240.1; -;
 DR EMBL; U95089; AAB53063.1; -;
 DR EMBL; U48722; AAC50802.1; -;

DR EMBL; U48723; AAC50804.1; -;
 DR EMBL; U48724; AAC50796.1; -;
 DR EMBL; U48725; AAC50797.1; -;
 DR EMBL; U48726; AAC50798.1; -;

Query Match 45.2%; Score 3099; DB 1; Length 1210;
 Beet Local Similarity 49.3%; Pred. No. 2,1e-158;
 Matches 626; Conservative 174; Mismatches 359; Indels 112; Gaps 23

11 LLLALLPPGAA--STVCTGTDMLRLPASPETHLMDRLRYGGCCVQVQGNLETLPTN 68
 14 LLALCPASALEEKVCQGSNTKLTQLGTFEDHPLSLQMFNNCEVVLNLEITVQRN 73
 69 ASLSFLQDIEVQGVYLIANQVRQVPLQRLIRVGTQLFEDNYALAVLDNGLNNTTP 128
 74 YLSFLKTIQEVAGYVLIANTVERIPLENLIQIRGMVYENSVALAVLSNYD----- 126
 129 VTGSPGGLRELQRLSTELLKGVLIQRNPOLCYDDTIIMKQIFKNNQALTLIDTR 188
 127 ---ANKTGLKELPERKLOETLHGAVRPSNNPALCNVESIQWRDIVSDPLSNMSPQNH 183
 189 SRACHPCSPMGKGRGSESEDCQSLTRTVCAAGCA-RCKGPLPTDCHEGCAACTGP 247
 184 LGSQCKDPCQPNQSCWAGAEBCQKLTIKICAQCCSGRGRGSPSDCHNQCACCTGP 243
 248 KHSQCLALFPHNSGICELHCPALVYNTDTPESMNPBGRTYFGASCYTACPYNYLSTD 307
 244 RESDCLVCRKFRDEATCKDCPPLMLYNPPTYQMDVNPBGKYSFGATCVKCKPRNVVTD 303
 308 VGSCTLVCPLHNOEVTAEEDTGRCCKSKPCARVCYGLQMFNNFTVSPMLRVKVASHL 367
 304 HGSQVRCACMADSYEM-EDQVRCCKCKBQPCRVNCGIGI-GEFKSL-----SIATNI 356
 368 EEPAGCKKIFGSLAFPEPSFDGPASNTAPQYIKANSKFIGITELGYLYISAMPDLP 427
 357 KHFNCTSIGDHLIIPVAPRGDSFHTPTLPDQELDIKTVKVEIGFLLIQMPENRND 416
 428 LSVQNLQVTRGRLINHGAYSLTLOGLSISMLGSLRLRELSGLALIHNTHCFTHTVP 487
 417 LHAENLEIRGRTRKQGGQFSIAVSLNTSLGRLSKETISQDVITISNKKLYCVANTIN 476
 488 MDLFRNPQALHTANRPEDCEVSGELACHQLCARHGCMGPRPTQVNCQGFRLQGEVC 547
 477 MKKLFGISGQKTIKINRGNSCKAIQVCHALCSPEGCMGEPKPCVSCRVNSRGREV 536
 548 BEGCVLQGLFREYVNAHCLPCHPCEQPNGSVTCFGRPADQCVACAHKDRPFCVACRP 607
 537 DKCNLLEGEPRFEVNSCEIQCHPECLPQAMNITCTGRPDNCCICAHYIDGPHCVKTCR 596
 608 SGVXPDLSTVPIKPPDEBGAQPCPCINCTHSCVDLDDKGCRAEQASLTIVSAVVS- 666
 597 AGVMGENTL-VKVIADAGHVCHLCHPNCCTYGGCTGEGECPTNGKRP--SIATGMVGA 653
 667 --LLVVVGVVEGILIKRROKIRKRYTRRLIOETLEVEPLTPSGAMPNQAKMILKET 724
 654 LLLLVVVALGIG---LFMRBRHIVRKTIRLLQREHEVLEPLTPSGEAPNQALLILKET 710
 725 ELKRVYLGSGAGCTYTKGIMPDGSNVKIPAIKYLRENTSPKANKELIDPAAVYAGVG 784
 711 EFKKIRVLGSGAGTYKGLMIPGSKVXI PAVKELREATSPKANKELIDPAAVYASVD 770
 785 SPVSRLLGLICTSTVQLTQMLPYCLLDHRENRGRIGSODLLNMCMQAKKGSYLED 844
 771 NPHVCCLLGICLSTVQLTQMLPFCCLLDYREHKNDIGSYLLNMCVQAKKNYLED 830
 845 VRLVHDLAARNVLVKSENVHKITDPGLARLLIDDETEYHADGKVPKIMWALSILRR 904
 831 RRLVHDLAARNVLTVPQHVITITFGAKLIGAEKEKTHAGKVPKIMWALSILHRI 890
 905 FTQSDVMSYGVTVMLTFGAKPPDGPAREIPOLLKEGRLPQPPICTDVYVIMWKC 964
 891 YTHQSDVMSYGVTVMLTFGSKPPDGPASISISILKGERLLPQPPICTDVYVIMWKC 950

```

QY 965 WMIDSECRPRELIVSEFSRMRARPORFVVIQ-NEDLPASPLDSTFYRSLDDEMDGL 1023
DB 951 WMIDADSRPKRELIIFFSKARBPQRYLVIQDERMHLSPFSDSNFYRALMDEEMDV 1010
QY 1024 VDAEYLVFQQGFPCPPAPAGAGKVMHRRSSSTRSGGDLTLGLEPSEEAARSLAP 1083
DB 1011 VDADEYLVFQQGF-----SSPSTRPLLS 1036
QY 1084 SEGASDVFDGDLGMAKGLQSLPTHDPSPLOKYSRDPVPLPSRT-DGYAPRLTSCP 1141
DB 1037 SLSATSN-NSTVACIRBNGLQSCPIKEDSFLOKYSDDPFGALTEDSIDTFL-----P 1088
QY 1142 QPEYVNGPVPARPPSPREGLPARPAGATLRAKTLSPKNGVYKDYAFAGAVENPE 1201
DB 1089 VPEYINQ-SVPKRPAGSVQNVYVNGPLN-----APSRDPHYQD--PHSTAVGNPE 1137
QY 1202 YL-TPGGGAAPQHPHPAPSPAPFNLYVWQ-----DP-----PERGAPSTRK 1244
DB 1138 YLNTVQ-----PFCVNSTFDSPAHMAQKSHQISLDNPDYQDFFPKAKRPNIGRK 1188
QY 1245 GTPTAENPEYL 1255
DB 1189 GS-TAENAEYL 1198

RESULT 5
EGRF_MOUSE STANDARD; PRT; 1210 AA.
ID EGRF_MOUSE
AC Q01279;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN EGRF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=liver;
RA MEDLINE=93026370; PubMed=1408137;
RA Avivi A., Skorecki K., Yavon A., Givol D.;
RT "Promoter region of the murine fibroblast growth factor receptor 2
RT (bek/KGFR) gene.";
RL Oncogene 7:1957-1962 (1992).
RN [2];
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c, and CD-1; TISSUE=liver, and Decidua;
RA MEDLINE=93126380; PubMed=7678348;
RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT "Expression of the epidermal growth factor receptor gene is regulated
RT in mouse blastocysts during delayed implantation.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59 (1993).
RN [3];
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=liver;
RA Hibbs M.L.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [4];
RP SEQUENCE FROM N.A.
RC STRAIN=B6/C3; TISSUE=liver;
RA MEDLINE=94170986; PubMed=8125255;
RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Eary H.S.,
RA Jenkins N.A., Lee D.C.;
RT "The mouse waved-2 phenotype results from a point mutation in the EGF
RT receptor tyrosine kinase.";
RL Genes Dev. 8:339-413 (1994).
RN [5];
RP SEQUENCE OF 1-714 FROM N.A.
RC TISSUE=Brain;
RA MEDLINE=91233866; PubMed=2030916;
RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;

```

```

RT "Comparison of EGF receptor sequences as a guide to study the ligand
RT binding site.";
RL Oncogene 6:673-676 (1991).
RN [6];
RP SEQUENCE OF 969-1117 FROM N.A.
RC STRAIN=C3H;
RA Eisenberg D.P., Serrero G.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X76987; CA55587.1; -
DR EMBL: U03425; AAA17899.1; -
DR EMBL: X59698; CAA4219.1; -
DR EMBL: L06864; AAA53029.1; -
DR EMBL: Z12608; CAA78249.1; -
DR HSSP: P11362; IFGK.
DR MED: MGI:95294; Egrf.
DR InterPro: IPR000494; EGRF_L domain.
DR InterPro: IPR000719; Euk_PKinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_PKinase.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L domain; 2.
DR ProDom: PD000001; Euk_PKinase; 1.
DR SMART: SM00261; FU; 3.
DR SMART: SM00219; TYRK; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane: Glycoprotein: Receptor; Signal: Transferase;
DR Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
RT SIGNAL 1 24
RT CHAIN 25 1210
RT DOMAIN 25 647
RT TRANSMEM 648 670
RT DOMAIN 671 1210
RT REPEAT 75 300
RT REPEAT 390 600
RT DOMAIN 1028 1071
RT NP_BIND 714 981
RT BINDING 720 728
RT ACT_SITE 747 747
RT ACT_SITE 839 839
RT DISULFID 190 199
RT DISULFID 194 207
RT DISULFID 215 223
RT DISULFID 219 231
RT DISULFID 232 240
RT DISULFID 236 248
RT DISULFID 251 260
RT DISULFID 264 291
RT DISULFID 295 307
RT DISULFID 311 326
RT DISULFID 329 333

```

```

FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 510 523 BY SIMILARITY.
FT DISULFID 526 535 BY SIMILARITY.
FT DISULFID 539 555 BY SIMILARITY.
FT DISULFID 558 571 BY SIMILARITY.
FT DISULFID 562 579 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 595 617 BY SIMILARITY.
FT DISULFID 620 628 BY SIMILARITY.
FT DISULFID 624 636 BY SIMILARITY.
FT MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 128 128 (POTENTIAL).
FT CARBOHYD 175 175 (POTENTIAL).
FT CARBOHYD 196 196 (POTENTIAL).
FT CARBOHYD 352 352 (POTENTIAL).
FT CARBOHYD 413 413 (POTENTIAL).
FT CARBOHYD 444 444 (POTENTIAL).
FT CARBOHYD 528 528 (POTENTIAL).
FT CARBOHYD 568 568 (POTENTIAL).
FT CARBOHYD 603 603 (POTENTIAL).
FT CARBOHYD 623 623 (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 4).
FT CONFLICT 1116 1116 HP -> DR (IN REF. 6).
SQ SEQUENCE 1210 AA, 134853 MW, 690E20D46DFD2F5 CRC64;

```

```

Query Match 44.9%; Score 3076; DB 1; Length 1210;
Best Local Similarity 49.2%; Pred. No. 3.6e-157;
Matches 630; Conservative 169; Mismatches 361; Indels 120; Gaps 26;

QY 11 LLLALLPPGAA--STQVCTGDMKLRPAASPTHLDLRHLYGQCVQVGNLELTLPFN 68
DB 14 LLLTALCAAGALBEKKKVCQGTNRNLQGLTFEDHFLSLQRMVNCVAVGNLEITVQGN 73
QY 69 ASLSFLDIOGVGVYLIANNOVQVPLQRLIVRGTOPEFNVAVALVDNGDPLNTPP 128
DB 74 YDSFLKTIQEVAVYVIALNTERIPLENLQIRGNALYENTYVALAISN----- 124
QY 129 VTGASPGGLRELQRLSTTEIKGVLIQRNPOLCYOTILMKCI-----PHKNQGLATLI 184
DB 125 -YGTNRIGLRLPMKRLQELILIGAVRPSNPILCMMDTITOMRDTIVONVMSNMDL--- 180
QY 185 DTNRSRACHPCSPWCKGSRGSSSEDCQSLRTVAGGA--RCKGRLPTDCHEQCAAG 243
DB 181 -QSHRSSCPKCDSPSCPGSCMGGEENCQKLTKICAOQCSHRCRGRSPSDCHQCAAG 239
QY 244 CTGPKHSDCLACJLHNHNSGICELHCPALTYNTDTESMPNPRGRTTFCASCYTACPNY 303
DB 240 CTGPRBBDCLVCKQFQDEATCKTCTPLMLYNTTQOMVNEGKYSFGATCVKCKPRNY 299
QY 304 LSTDVSGCTLVCELHNOEVAEDGTORCEKSKPCARVCYGLGM--FNNFTVSFMLRVK 361
DB 300 VVTDHSGCVACGPDYEV--EEDGIRKCKKCDGPKKVCNGIGIGFHK--TIS----- 350
QY 362 VSASHLEEFAGCKKTFQSLAFLEESDGDPASTAAQYKANKSKFTIGTELTVYISAM 421
DB 351 INATNIGHFYKCAISGDHILVAFKGSFTPTPLDRELEILKTYEITIGFLLIQAM 410
QY 422 PSLPLDLVFPQNTQVIRGLIHNGAVSLTQGLISMLGRSLRELGSGLAIHNTILC 481
DB 411 PDNMTDLHAFENLEIRGRTKOHCQSLAVGLNITSLGRSLKELSDGVYISGRNVL 470
QY 482 FVATVMDQLFRPHQALLHTANRPDEVCVGGSLACHQLCARHGCMWGPPTQCVNCSOL 541
DB 471 YANTINMKLFGFPNOKTKIMNRAEKDKAVNHNPLCSSEGCWGPBRDVCVCSQVUS 530
QY 542 RGQECVEGCVLQGLPREYVNAHCLPCHREQOPQNGSVTCGPRADQCVACAHKDPFP 601

```

```

DB 531 RGRCEKCNILLEGEPREPVNSBICQCHPECLPQANNTCTGRGPDNCTIQCAHYIDGPH 590
QY 602 CVACPSGVNEDLSYMPITWKFPRDEGACQCCPINCSTSCYDLDKCGPACRASPILTSIV 661
DB 591 CVKCPAGIMGNNTL-VMKYAADANNVCHLCHANKCTYGGAGPGLQCGEVPWSPGPKPISA 645
QY 662 SAVGILLVVLGVVFGI-LIKRQKIRKTYRRLLOETELVEPTLPSGAMNQAMRI 720
DB 650 TGIYGLLFTIV-VALGILGPMRRRIIVRRITRLRLOESELVEPTLPSGARNOAHRI 708
QY 721 LKTELKRVKVLGSGAFGYIKGIVPDGENVKI PVAIKYLRNTPSPKANKELIDEAAYW 780
DB 709 LKTEFPIKVLGSGAGTYVKGIMPECEKVKI PVAIKELRATSPKANKELIDEAAYW 768
QY 781 AGVSPVSRLLGICLSTVQVLTQCLMPVCCILDHRENRGRIGSODLIMWCQIAKMS 840
DB 769 ASVNPVHCRLICLSTVQVLTQCLMPVCCILDHRENRGRIGSODLIMWCQIAKMS 828
QY 841 YLEDRVLVHRDLAARVLYKSPNHVXITDFGLARLLDIDETEHADGKVPIMMALESI 900
DB 829 YLEDRVLVHRDLAARVLYKTPQHVKITDFGLAKLGAEEKEVHAEGKVPIMMALESI 888
QY 901 LREFTQSDVMSGYVYVWELMTFGAKPYGIPARELPDLEKGERLPQPPITIDVYMI 960
DB 889 LHRITQSDVMSGYVYVWELMTFGSPYDIPASDISISILEKGERLPQPPITIDVYMI 948
QY 961 MVKWMIMDSRPFRELVSEFSRMARDPQRFVIO-NEPLGASPLDSTFYSLEDDO 101
DB 949 MVKWMIMDASRPFRFLLIEFSKARDPQRYVIVIGDERMHLPSPTDSNFRALNDEED 100
QY 1020 MGDVDAEVLVPOQGFPCPDPAAGAGVNHHRSSSTRSGGDLTLGLESEEBAPRS 107
DB 1009 MEDVDADEVILVPOQGF-----NSPST-----SRT 103
QY 1080 PLASBEGASDVDDGDLGMAAGLOSPLTHDSPLQKRSSEPTVLPSET--DGIVAPL 113
DB 1035 PLUSSLSATSN---NSTVACINRNGSCRYKEDAFQKQYSSPTGAVTBDNIDDAFL--- 108
QY 1138 TCSPOPEYVNPQVVRPQPSPREGRPLPAARPAAGATLERAKTISPQNGGVKVFAGFAGV 119
DB 1088 ---PVPEYVNG-SVPRKPAQSGVNPVYHNGPLHP-----APGDLHYQN--PHSNV 113
QY 1198 ENPEYLT-TPQGAAPQPHPPAPSPAFNDLWYDQ-----DF-----PERGAP 124
DB 1134 GNPXYLNTAQ-----PTLSSFSNPALMIQSHQMSLDNDYQDDFFPKETKN 118
QY 1241 STFGKPTAENPEYLG-DVY 1260
DB 1185 GIKRG-PTAENAEYLKVARP 1203

```

RESULT 6
 ERB4_HUMAN STANDARD; PRT; 1308 AA.
 AC Q15303;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
 DE (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
 GN ERB4 OR HER4.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM JM-A).
 RC TISSUE=Breast carcinoma;
 RX MEDLINE=93189574; PubMed=839326;
 RA Plozman G.D., Cuioscu J.-M., Whitney G.S., Green J.M., Carlton G.W.,
 RA Roy L., Neudauer M.G., Shoyab M.;
 "ligand-specific activation of HER4/p180erbB4, a fourth member of the

RT epidermal growth factor receptor family.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
 RC TISSUE=Fetal Brain;
 RX MEDLINE=97476287; PubMed=9334263;
 RA Elenius K., Cortas G., Paul S., Choi C.J., Rio C., Ploman G.D.,
 RA Klagsbrun M.;
 RT "A novel juxtamembrane domain isoform of HER4/erbB4. Isoform-specific
 RT tissue distribution and differential processing in response to
 RT phorbol ester";
 RL J. Biol. Chem. 272:26761-26766(1997).
 CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
 CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
 CC NTRK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
 CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
 CC RECEPTORS (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
 CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
 CC BUT NOT JM-B. SO THEY RESPECTIVELY REPRESENT CLEAVABLE AND
 CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
 CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, CEREBELLUM,
 CC KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,
 CC PITUITARY, SPLEEN, TESTIS AND PANCREAS. LOWER LEVELS IN THYMUS,
 CC LUNG, SALIVARY GLAND, AND PANCREAS.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL databank.
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: L07668; AAB59446.1; -;
 DR HSSP: P11362; JPKG.
 DR Genew; HGNC:3432; ERBB4.
 DR MIM; 600543; -;
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_kinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_kinase.
 DR InterPro; IPR004019; YLP_motif.
 DR Pfam; PF00069; Kinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR Pfam; PF02757; YLP; 2.
 DR ProDom; PD000001; Euk_kinase; 1.
 DR SMART; SM00261; FU; 4.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KM Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KM Alternative splicing.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
 FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 652 675 POTENTIAL.
 FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 186 334 CYS-RICH.
 FT DOMAIN 496 633 CYS-RICH.
 FT DOMAIN 718 985 PROTEIN KINASE.

FT NP_BIND 724 732 ATP (BY SIMILARITY).
 FT BINDING 751 ATP (BY SIMILARITY).
 FT ACT_SITE 843 843 BY SIMILARITY.
 FT DISULFID 189 197 BY SIMILARITY.
 FT DISULFID 193 205 BY SIMILARITY.
 FT DISULFID 213 221 BY SIMILARITY.
 FT DISULFID 217 229 BY SIMILARITY.
 FT DISULFID 230 238 BY SIMILARITY.
 FT DISULFID 234 246 BY SIMILARITY.
 FT DISULFID 249 258 BY SIMILARITY.
 FT DISULFID 262 289 BY SIMILARITY.
 FT DISULFID 293 304 BY SIMILARITY.
 FT DISULFID 308 323 BY SIMILARITY.
 FT DISULFID 326 330 BY SIMILARITY.
 FT DISULFID 503 512 BY SIMILARITY.
 FT DISULFID 507 520 BY SIMILARITY.
 FT DISULFID 523 532 BY SIMILARITY.
 FT DISULFID 536 552 BY SIMILARITY.
 FT DISULFID 555 569 BY SIMILARITY.
 FT DISULFID 559 577 BY SIMILARITY.
 FT DISULFID 580 589 BY SIMILARITY.
 FT DISULFID 593 614 BY SIMILARITY.
 FT DISULFID 617 625 BY SIMILARITY.
 FT DISULFID 621 633 BY SIMILARITY.
 FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 138 138 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 174 174 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 181 181 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 253 253 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 358 358 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 473 473 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 548 548 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 620 620 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARSPLIC 626 648 NGPISHCIIYPTMGHGLPQHA -> IGSSIEDCLGLMD
 FT (IN ISOFORM JM-B).
 SQ SEQUENCE 1308 AA; 146807 MW; 5FA8B0965D86761 C6C64;
 Query Match 42.8%; Score 2934.5; DB 1; Length 1308;
 Best Local Similarity 44.7%; Pred. No. 1,4e-149;
 Matches 605; Conservative 188; Mismatches 380; Indels 179; Gaps 29
 QY 9 MGLLALIPGAA-----STQVCTGTOMKRLPSPETHLDMLEHLYOGGVGVGNEELY 64
 DB 8 WWSVSLVAACTVOPSPDSQSVCACTEKLSSLDLQGYRALRKRYENGEVVGNETITS 67
 QY 65 LPTNASLFDLDIOEVGVYLIANQVQVPLORLIRVGTQLFEDNYALAVLDNDPILN 124
 DB 68 IEHNRDLSFLRSVSEVGVYLVNLNGRYVPLENLRIRKTKYEDRYALALFLNVRKXG 127
 QY 125 NTTPTVTAASPGALRELDRLSLTEILKSGVLIQNNPQLCYDITLMDITRKNQALLTLI 184
 DB 128 NF-----GLQELGLKNLITLNGGVYVDNKKLCYADITHMDIVRNPPSVNLTLV 178
 QY 185 DTNRSRACHPCSPKCKSRRCMGESSEDCSLRTVYAGGC-ARCKGFLPLPDCCHQCCAG 243
 DB 179 STNASSCGGRCHSKCTG-RMGPTENHHCQLITTVACBQCDDGCGYPPYSDCHRCACG 237
 QY 244 CTGRHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
 DB 238 CSGKDDCDCAKMFNDGACVQCQGTFFVYNTFTQLBHNFAKTYGAFQYCKCPHNF 297
 QY 304 LSTVGSCTLVCPPLHNGEVTAEADGTORCECKSPCARVYCGYGMFNNFTVSMRLVRPKYS 363
 DB 298 V-VDSSCVACFSSSKREV-EEGIMKCKCTDTCFKAQDGI-----TGLMSAQVTD 349
 QY 364 ASHLEEFAGGKKITFGSLALPESFDDPASNAPQYIKANSKFTIGITELTGLVLYISAMPD 423

```

Db 350 SSNIDKFINCTKINGNLFIIVTGIGHDPYNAIEAIDBEKLNVEFTVEITGTINIQSWPE 409
Qy 424 SLPLDSVFQNLQVIRGRILHNGAVSLTLQIGISMLGRSLREISGLALIHNTHLCPV 483
Db 410 NMTDFSVSNLVTIGRVLISGLSLILIKQIGITSLPQSIKELISANITYITNSNLCTY 469
Qy 484 HTVPMDOLFNRPHQALLHTANRPREDECEVGEGLACHOLACRCHCMGPGPTOCVACSGFLR 543
Db 470 HTIMWTLFTSTINGIIVIRDRKAKENCTAEGMVCNHLCSDDGCGPGPDCLSCRFPSRC 529
Qy 544 QECVECEVLOGLPREVYNACHLPCHECOP-ONGSVTCFPGPADQCVACATYKPPFC 602
Db 530 RICESCNLYGGEPEFENGSIYCECPQCEKMDGLTCHGPPDNCCTKCSHFKGPPNC 569
Qy 603 VARCPGVKPDLSYMPIMKFPDEGACPCPCINCTHSCVDLDDKGC-----PAE 651
Db 590 VEKCPDGLQAGNSF--IFKVADPDECHCPHNPCTQGCNPTSHDCIYYPWTHSTLPPH 647
Qy 652 QRASPLTIVSAV- GILLVVVLGVVGIILIKRQOKIRKYTKRRLLOETELVEPLTPSG 710
Db 648 AR-TPL--IAAGVIGGLFIVIGLTFVAVYRRKSIK-KKRALRFL-ETELVEPLTPSG 702
Qy 711 AMPNOAKRILKETELRKVKYLGSGAGFTYKGIWIDGENVKIPVATKLENTSPKAN 770
Db 703 TAPNOAQLRIKETELRKVKYLGSGAGFTYKGIWIDGENVKIPVATKLENTSPKAN 762
Qy 771 KEILDEAYMAGVGSFYVSRLLGICLTSTVOLVTLQMPYGLLDHVENRGRIGSDPLLN 830
Db 763 VEPMDALIMASMDPHVLVRLIGVCLSTPTQLVTLQMPHGLCLEYVHEKDNIGSQLLN 822
Qy 831 WMOJAKGMSYLEDVRLVHRDLAARNVLYKSPNVKLTDRGLAALDIDETVHAOGKY 850
Db 823 WCVQIAKGMVLEERLVRDLAARNVLYKSPNVKLTDRGLAALDIDETVHAOGKY 862
Qy 891 PIKMALESILRRPETHQSDVMSYGVTWELMTGCAPIYDGIPIREIPDLLEKGERLP 950
Db 883 PIKMALESILRRPETHQSDVMSYGVTWELMTGCAPIYDGIPIREIPDLLEKGERLP 942
Qy 951 PICTIDVTMVKCMIDSECRPRFRELVSFSSMADPQRFVYIQNE- LGPASPLDST 1009
Db 943 PICTIDVTMVKCMIDSECRPRFRELVSFSSMADPQRFVYIQNE- LGPASPLDST 1002
Qy 1010 FYRSLLEDDMDGLVDABEYLVPQGFPCPPAPAGAGMWHRRSSSTRSGGDLTLGL 1069
Db 1003 FQNLDEEDLEDMMDABEYLVP-QARNIPTP-----ITSRKIDISNS-----EIGH 1050
Qy 1070 EPSEEBARS-----PLAP-SEGASDVFPDGLMG 1100
Db 1051 SPPPAYTPMSGNQFYVBDGFAEGVSVPRAPTSTIPEAPVAGATAEIPDSCNGT 1110
Qy 1101 AKGLOSLFTHDPSPIQRYSEDPVPLPS-----ETDGVAPLTCSPQREYNQDVP 1153
Db 1111 LRKPVAVHVEDSSTQRYSDPTFAPAPRSRGGLDEGVTTPRDKPKKEYLNPVE-- 1167
Qy 1154 QPPSPREPLPAPAPAGATLERAKTLSPGKXGVKDVAFAGAVENPEYLPQGAAPQ 1213
Db 1168 -----ENPFVSR-----KNGDQ-----ALDNEFYNAANG----- 1194
Qy 1214 HPPPA-----FSPAFDNTYYWDQPPERGA- 1238
Db 1195 -PPKADEYVNEPLYLNTFANTLGAEYLNKXNLSMEKAKKAFDNDVYMHSLPPRSTL 1253
Qy 1239 -PPSTFKGTP-----AENPEYL 1255
Db 1254 QHPDYLOEYSTKYFYKONGRIRPIVAENPEYL 1285

```

```

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
CN ERB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Balliga R.R., Opel D.J., Han X.,
RT "Neuregulins promote survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and Erb4 expression in neonatal and adult
RT ventricular myocytes."
RL J. Biol. Chem. 273:10261-10269 (1998).
RN (2)
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system."
RL Neuron 6:691-704 (1991).
RN (3)
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration."
RL J. Neurosci. 17:1642-1659 (1997).
CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTRK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF- $\alpha$ , AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC The SWISS-PROT entry is copyright. It is produced through a collaborative
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF041838; AAD08899.1; -.
CC EMBL: U52531; AAC53051.1; -.
CC HSSP: P11362; IFCK.
CC InterPro: IPR000494; EGFR_L domain.
CC InterPro: IPR000719; Euk_Pkinase.
CC InterPro: IPR002174; Furin-like.
CC InterPro: IPR001245; Tyr_Pkinase.
CC InterPro: IPR004019; YLP motif.
CC Pfam: PF00757; Furin-like; 1.
CC Pfam: PF00069; Pkinase; 1.
CC Pfam: PF01030; Recep_L domain; 2.
CC Pfam: PF02757; YLP; 2.
CC PRINTS: PR00109; TYRKINASE.
CC ProDom: PD000001; Euk_Pkinase; 1.

```



```

Db      1195 --PKAEDEVNEBLYINTFNALGNNAEMKNSLSVPEKAKKAFNDPDYVNHSLPPRST 1252
QY      1239 --PSTFKGFT-----AENPEYL 1255
Db      1253 LQHPDYLOEYSTKIFYKONGRIRPIVANPEBYL 1285

RESULT 8
XMRK_XIPMA
ID_XMRK_XIPMA STANDARD; PRT; 1167 AA.
AC P13388:
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
GN XMRK OR TU.
OS Xiphophorus maculatus (Southern platyfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_TaxID=8083;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90015140; PubMed=2797166;
RA Wiltbrodt J., Adam D., Malitschek B., Mauelel W., Raulf F.,
RA Telling A., Robertson S.M., Schachtl M.;
RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
RT inducing Tu locus in Xiphophorus.";
RL Nature 341:415-421(1998).
RN [2]
RP REVISION TO 515.
RA Schachtl M.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL, constitution
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X16891; CAA34770.2; -
DR PIR; S06142; S06142.
DR HSSP; P13362; 1R6K.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_Thr_Pkinase.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Reced_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TYKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR TRANSMEMBRANE; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
FT SIGNAL 1 25
FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE

```

```

FT DOMAIN 26 642 KINASE.
FT TRANSMEM 643 665 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 666 1167 POTENTIAL.
FT DOMAIN 710 977 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 716 724 PROTEIN KINASE.
FT BINDING 743 743 ATP (BY SIMILARITY).
FT ACT_SITE 835 835 BY SIMILARITY.
FT DISULFID 195 204 BY SIMILARITY.
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 220 228 BY SIMILARITY.
FT DISULFID 224 236 BY SIMILARITY.
FT DISULFID 237 245 BY SIMILARITY.
FT DISULFID 241 253 BY SIMILARITY.
FT DISULFID 256 265 BY SIMILARITY.
FT DISULFID 269 296 BY SIMILARITY.
FT DISULFID 300 311 BY SIMILARITY.
FT DISULFID 315 330 BY SIMILARITY.
FT DISULFID 333 337 BY SIMILARITY.
FT DISULFID 504 513 BY SIMILARITY.
FT DISULFID 508 521 BY SIMILARITY.
FT DISULFID 524 533 BY SIMILARITY.
FT DISULFID 537 553 BY SIMILARITY.
FT DISULFID 556 569 BY SIMILARITY.
FT DISULFID 560 577 BY SIMILARITY.
FT DISULFID 593 615 BY SIMILARITY.
FT DISULFID 618 626 BY SIMILARITY.
FT CARBOHYD 622 634 BY SIMILARITY.
FT CARBOHYD 614 114 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 144 144 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 201 201 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 356 356 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 365 365 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 398 398 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 621 621 N-LINKED (GLCNAC. .) (POTENTIAL).
FT SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;

Query Match 38.7%; Score 2655.5; DB 1; Length 1167;
Best Local Similarity 44.6%; Pred. No. 1e-134;
Matches 569; Conservative 173; Mismatches 384; Indels 151; Gaps 31

QY 4 AALTRGGLLALPEPGAST---QVCTGDMKRLRPASPTHLDMLRHLYGCGVYQGN 59
DB 8 AALLQ--LLVLSTISRCSTDPDKVCOGRTSNQMTM---LNNHYLMKKKMYSGCNVLEN 62
QY 60 LELTYLPTNASLSTLDIOGVGVVLIHNVQVQVPLQRLIRIGRQLPFDNALAVLDN 119
DB 63 LELTYTOENODLSFLQSIQEVGGVLIAMNEVSTIPLVRLRLRGONLYEGNFTLLVMSN 122
QY 120 GDPUNNTPTVYGASPGGLRELRSLTEILKGVLIQRNFDLCYODTILMKDIFHNQNL 179
DB 123 YQK-NPSSP--DVYQVGLKQLQSLNTEILISGVKVSHPNLLCNVETIMWMDIVDTSTNP 179
QY 180 ALTLIDTRSRBACHPCSPMKSGRCWGESSEDDQS.LTRYCAGGC-ARCKGPIPTCCGE 238
DB 180 TWNLIIPAFERQCKCHGCVNGSSCAVPGHGHQKPTKLCAEQCNRRRGPPYDDCNE 239
QY 239 QCAAGCTGPKHSDCLAFHFNHSGICELHCPALVTYNTDFESMPPEERYFGASCVTA 298
DB 240 HCAGCGCTGPRATDCLAKRDENDGTCKDTCPPKIKYDIYSHQVNDNPNIKYTFGAACVGE 299
QY 299 CPNYYLSTDVGSCTLVCPHNLQVTAEDGTQRCCKSPCARCYGLGMFN-NFTVSPFL 357
DB 300 CPNYYVYTE-GACVRSASGMLVD-ENGKRSCKPCDGVCPKCDGIGISLNTIA--- 354
QY 358 RVPVSAHLEEFAGCKIFGSLAPLPESRGGP---ASNTAPQYIKANSKPIGITEILNG 414
DB 355 ---VNSTINRSNCKTKINGDIILNRNSFEGDPHYKIGTMDPEHL---WNLTTYKEITG 407
QY 415 YLYISAMPDSLPLSVFQNLQVIRGRILHNGAVS-LTLQGLGISWLGRLSRLREGSLAL 473

```



```

01 307 DVSGCTLVGCHLHNGQETLADGTQOREBKSGKPARVCYGLGMPNNFYLSFMLRVKVSASH 361
02 Db 297 DQTS CVAAQCPDRKEDV-KNGLKKEPCGGCLCPKACEGTGSGSRFQT-----VSSN 341
03 367 LEEFAGCKKIFGSLAFLEPSPFGDPASTAPQYIYANSKF:GIFELTGCLYISAMPDLP 421
04 QY 348 IDGVNCTKILGNLDLITGLNGDPMHKIPALDPRKLVNFTVAIEITQYLNIGMPHHM 401
05 Db 427 DLSVFQNLVIRGRILHNGAYS-LTLQGLGISWGLSRLFELSGLALIHNNTHLCEVHT 481
06 QY 408 NFSVFNSLTLTIGGRSLYNGFSLILMKYLNMTSLGFRSLKHSAGRIYISANROLCHHS 461
07 Db 486 VPMDOLEFRNHQALHTA-NRPEDCSVEGLACHQGLCARGCHWCBGPFCQCNCGQFLRG 544
08 QY 468 LMTKTLRGTERLRDILKNNRPRDCCVAEGKCDPLSGGCGMGPBGQULSCNNYSRG 521
09 Db 545 ECVEECRVLOGLPREVYNAHCLPCHPECPONSGSVTCFGEADQVACAHYKDPCEVA 601
10 QY 528 VCVTHCNFLNGEPRFEAEHAEFCFSCHPECPMEGTATCNGSGSDTCAQCAHFRDGHCVS 581
11 Db 605 RCQSGVXKPDLSYPMKPRDEBGAQOPPCINCHTSQVUDLDKXGPAQORA----SPLT 661
12 QY 588 SCFPGVLG--AKGPIKYKPDVONECRPHENCTOGCGPELQDLQGLVLVIGKHTLMA 641
13 Db 661 VSAVVGILLVVLVGVFGILLIKRQOKIR-KYTWRLLOETELVEPLTPSGAMPNQAOMR 711
14 QY 646 LTVIAG--LVVIFMVLGCFELWRRRIQNNBRARVIREBESIPLDPS-EKANKYLAR 702
15 Db 720 ILKETELRKVKVLGSCAFGTQYKGIIMDGENVKI PAIKYLRENTSFKANKETLDEAYV 779
16 QY 703 IFKETELRKLVIGSGSVFGTVKGVWIEGESIKIPVCIKYIEBKSRGOSFOAATDHMLA 761
17 Db 780 MACVGPYVSRLLIGLCTETVOLQNLMPYCGLLDHYRENNGRIGSODLLMNCQJAKGM 831
18 QY 763 IGLDHAHIVRLILGCPGSSDQVLYQYPLQSLDHRQNHGALGPQLLNMVQJAKGM 822
19 Db 840 SYLEDELVLRDLAARNVLVSPNHVKITDFGLARLLIDETEHADGKVP:KMMALES 891
20 QY 823 YVLEEGMWHNRLLAARNVLLKSPQOVADGVADLLPRDDKOLLYEAKTPIKMMALES 882
21 Db 900 ILARRFTHQSDVWASGVTWELMTFGAKPYDGIARERLIDLEKGERLQPRPCTIDVYM 951
22 QY 883 IHRGKTIHQSDVMSGVTVWELMTFGARPYGLARLAEPRDLLEKGERLAQPOLCTIDVYM 942
23 Db 960 IYKCMWIDSECRPRERELVEFFSFMADPPARFVIONEDLGPV---SPLDSTFYRSLLE 1011
24 QY 943 VYKCMWIDENIRPTFKELANEFTHMADPRVLYIKRES-GPEIARGPBRHGLTKXLE 1001
25 Db 1017 DDIMGGLVVAEEVLPQCGFFCPRDPAFAGAGMHHRRSSSTRSGGDDLTLGLEP-SEEE 1071
26 QY 1002 EYVLEDELIDLDLEAED-----NATTLTIGSALSLPVGTUNRR 1041
27 Db 1076 APRSP:LAPESEAGSDVDFGDLQMGAKLQSLPHND-PSPLQRYSEDVTVLP-----S 1121
28 QY 1043 GSGSLTSPSSGV-MMMNGNGLGSEQGEABVAGSSERCEPRVSLH-----PMRGCLASE 1091
29 Db 1129 ETDGYIA-----PLTSGPQPE---YVNOQDVRPOPSPSREGP----- 1161
30 QY 1096 SSEGHTYGSSEALQEKVMSCRSRSRSPRPRGDSAVHSORHSLITPTVPLSPGLEED 1151
31 Db 1163 -----LPARPAAGATLLEAKTLPSP-GKXGVV-----KDVAFAGGANENREYVLP 1211
32 QY 1156 VNGYVMPDLHLKGTSSREGTILSSVGLSSVLGTEEDED-----ELEYNNRRRRHS 1201
33 Db 1211 PQCHPPAPASPAFDNLVYWD-----QDPRRGAPPSTFKQTPAENPEYL 1251
34 QY 1208 P-PHPRPSPSLEBELGEYEMDVGSDLSASLGSTQCSLHPVPIWPTAGTTPDEDEYEM 1263

```

DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3).
GN ERB3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=96096535; PubMed=8522190;
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
RT "Cloning of the rat Erb3 cDNA and characterization of the
RT recombinant protein".
RL Gene 165:279-284(1995).
RN [2]
RP REVISIONS TO 85; 513 AND 565.
RA Hellyer N.J., Koland J.G.;
RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE OF 922-1097 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=sciatic nerve;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Froment P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuroregulin and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration".
RL J. Neurosci. 17:1642-1659(1997).
CC - FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC - SUBCELLULAR LOCATION: Type I membrane protein.
CC - DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE.
CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib.ch).
CC -----
DR EMBL: U29319; AAC38498.2; -
DR EMBL: U52530; AAC3050.1; -
DR HSSP: F11362; IFGK.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam: PR00069; kinase.1.
DR Pfam: PR00757; Furin-like.1.
DR Pfam: PR01030; Recep_L_domain.2.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_kinase.1.
DR SMART: SMO0261; FU; 5.
DR SMART: SMO0219; TyrcK.1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP_FALSE_NEG.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR.1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM.1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 19
FT CHAIN 20 1339 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT DOMAIN 20 643 EXTRACELLULAR (POTENTIAL).
FT

FT TRANSMEM 644 662 POTENTIAL.
FT DOMAIN 663 1339 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 183 259 CYS-RICH.
FT DOMAIN 707 964 PROTEIN KINASE.
FT NP_BIND 713 721 ATP (BY SIMILARITY).
FT BINDING 740 740 ATP (BY SIMILARITY).
FT ACT_SITE 832 832 BY SIMILARITY.
FT DISULFID 186 194 BY SIMILARITY.
FT DISULFID 190 202 BY SIMILARITY.
FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.
FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 266 BY SIMILARITY.
FT DISULFID 280 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 333 353 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC.) (POTENTIAL).
FT CONFLICT 108 108 L -> P (IN REF. 3).
SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402BDF0F1E CRC64;

Query Match 33.8%; Score 2317.5; DB 1; Length 1339;
Best Local Similarity 40.2%; Pred. No. 1,4e-116;
Matches 518; Conservative 176; Mismatches 434; Indels 161; Caps 34

QY 3 LAALCRWGLLALLPPGAA---STQVCTGDMKRLRPASETHLDMLRHLYGCGOVVQGN 59
DB 7 LQVLC-----FLSLARGSEMNSQAVCPGTLNGSLVGDADNGVQTLKYKCEVVMGN 62
QY 60 LEITLPTNASTFLQDIQEVQVLLAHNVQVPLQRLIRYRGTOLEFDNYALAVLDN 115
DB 63 LEIVTGHNADSLFLQWIREVTGYLVAMNEFSVPLPNLRVVRGVGVYDGKFAIFVM-- 120
QY 120 GDPINNTTPVTGASPGGIRELQSLTEILKGVLLQRPOLQYQDTILMKDIFKHNQL 179
DB 121 ---LNYNT-----NSSHALRQKLFQTLREILISGGVYIEKNKLCHEMTIDRIVRR-- 170
QY 180 ALTLIDTRSPACIPCPGPMCKGSCWSESSEDCOSLTRVCAGGC-ARCGPLPTDCHE 238
DB 171 GAETIVKNGANGACPPCEHCXG-RCWEPGRDDCQILTKTICAQCNCRGCPMPNQCCHD 229
QY 239 QCAAGCTPKKSDCLACILFHNSGICGLHPALVTYVTDTFESMPNPEGRVYTGASCVTA 228
DB 230 ECAGGCSPPQDTDFACRRPNDGACVPRCPBPVYTKLTFQLEPNPHTRYQGVGVAS 289
QY 299 CPVNYLSTDVSGCTLVCPPLHNQEVYTAJDGTQRCCKSKPCARVCYGIQMFNFTVSFWLR 358
DB 290 CPHNFV-VDQTFVCVRACPDPKMEVD-EGHGLKMEBPCCGLCPKXACEGSGS-----R 339
QY 359 VPKVSASHLEFACCKIFGSLAFPLSPGDPDASNTAPQYIKANSKFTIGITELNGYLYI 418
DB 340 YQYDSSNIQDFVNYCTKLIGLNDPLITGLAVDPMHKIPALDPKLVNFTVREITGYLNI 399
QY 419 SAMPDSLPDLSVFQNLQVIRGRILHNGAVS-LTLQGLGISWLGRLSRLRELQGLALIHNN 477

```

Db      400 QSWPPHNHNFSVSNLTITIGRSLYNRGSLIMKLANLTSIGFSLKEISAGRYISAN 459
Qy      478 THLCFYHVPMDLFRPHQALLHTA-NRPEDCEVEGLAHCQLCARHGCMKPGTQVCN 536
Db      460 QQLCYHSHLMTLLKSPSEERLDITKRLDGLGLBGRVCDPLSGGGCWPGGQGLS 519
Qy      537 CSQFLRGQECVECECVLYGLPREVYNAHRLCPHEPCOPONGSVTCFGEADQVCACAHY 596
Db      520 CRNYSREGVCTVHCNPLQGEPRFEVHAQCFSCHPECLPMEGSTONGSGDACARCAHF 579
Qy      597 KDPFCVARGCPSPGKPDLSYPIWKFPDEGACQCPINICTSC--VLDLDDGCAEORA 654
Db      580 RDGPHCVNSCPHGLG--AKGPIYKYPDQNECRPCHENCTGCGNPELDLQGAELVM 637
Qy      655 SPLTSIVSAVIGILLVVLGVGVGILIKRQOKIR-KYMRRLLOTELVEPLTSPGAMP 713
Db      638 SKPHLVIAVVG--LAVILMLIGSGFLYWRGRIRQKAMRMYLERGESIEFLDLS-EKA 694
Qy      714 NQAMRLKETELRKVKYUGSAGFVYVYGIWPDGENVKIVAIKVLRENTSPANKET 773
Db      695 NKYLARLFKTELKRLKVLGSGVGVTHKGIMPEGESIKIVICKIYEDKSGRSFOAV 754
Qy      774 LDEAYVAVAGVSPYVRLLGICLTSTVQVLTOLMPYGCLLDHYRENRGLSGODLLNMC 833
Db      755 TDHMLAVGSLDHAHYRLGLCGSSLLQVLTQYLPGLSLDHVKQRETLGQDLNMGV 814
Qy      834 QIAKMSYLEDELVHRDLAARNVLYKSPNNHYKITDFGLARLLDIDETEYHAGGKVPIC 893
Db      815 QIAKMWYLEEHSVMVRDLALRNVMKLSPSQVADFGVADLLPDDQLLHSEAKPTIK 874
Qy      894 WMALSTLRRTFHOSDWSYGVTVWELMTFGAKPFDGIPAEIPIDLLEKGRLEPQPTIC 953
Db      875 WMALSTHFGKYHOSDWSYGVTVWELMTFGAEPAGRLAEIDLEKGRRLAQPIC 934
Qy      954 TIDVIMVYKMMIDSECPREFELTSEFSRMAPDQRFVVIQNEIDLGPASPLDSTFYS 1013
Db      935 TIDVIMVYKMMIDNINPTFEKLANETFRAPADPRRLVTKRAS-GRGTP--PAAPS 991
Qy      1014 LLEDDMGDLVDAEELVLPQCGFCFDPAPAGAMVHRHRSSSTRSGGGLITLLEPSE 1073
Db      992 VLTTEL--QEALEPEL-----DLDLLEAE 1017
Qy      1074 E-----EAPRSPLAPSEG-----AGSDVPDGLGMAKALQ 1105
Db      1018 EGLATLGLSALPTCTTTRPGSQGLSPSSGYPMNMOSSIGECIDSAVYLGHEQFSR 1077
Qy      1106 SLPTHDPSPLOQYSEDPVPLDSEITDGYV---APL-----TC-----SPOE---Y 1145
Db      1078 PLSLH-PIPRGR-----PASESSCHVTGSEAELEKVSVCRSRSRSPRPGDSAY 1129
Qy      1146 VNGPVRPQPSPREP-----LPARPGATLERAKTILSP-GKNGV----- 1187
Db      1130 HSORHSLPLVPLSPGLEEDNGYVDPDHLRGASSREGTILSVGLTSGTEBED 1189
Qy      1188 KDVFARFAGAVENPEYLLTPOGAAPQPHPP 1216
Db      1190 ED-----EEYEVNKRKRGRSP-PRPP 1209

```

```

OC      Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX      NCBI_TaxID=7227;
RN      [1]
RP      SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX      MEDLINE=94350209; PubMed=8070664;
RA      Clifford R., Schubach T.;
RT      "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT      that several genetically defined classes of alleles cluster in
RT      subdomains of the receptor protein.";
RL      Genetics 137:531-550(1994).
RN      [2]
RP      REVISIONS.
RA      Clifford R., Schubach T.;
RL      Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX      MEDLINE=85124611; PubMed=2992499;
RA      Liven E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT      "The Drosophila EGF receptor gene homolog: conservation of both
RT      hormone binding and kinase domains.";
RL      Cell 40:599-607(1985).
RN      [4]
RP      SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RX      MEDLINE=Oeagon-R; TISSUE=Embryo;
RA      MEDLINE=87002474; PubMed=3093080;
RT      Schlegler E.D., Segal D., Glazer L., Shilo B.-Z.;
RT      "Alternative 5' exons and tissue-specific expression of the
RT      Drosophila EGF receptor homolog transcripts.";
RL      Cell 46:1091-1101(1996).
RN      [5]
RP      SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP      ANALYSIS.
RX      MEDLINE=99102120; PubMed=9882502;
RA      Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT      "Several levels of EGF receptor signaling during photoreceptor
RT      specification in wild-type, Ellipse, and null mutant Drosophila.";
RL      Dev. Biol. 205:1129-1144(1999).
RN      [6]
RP      SEQUENCE FROM N.A. (ISOFORM TYPE I).
RX      STRAIN=Beckley;
RA      MEDLINE=20196006; PubMed=10731132;
RA      Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA      Anagnostis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA      George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA      Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA      Branton R.C., Rogers Y.-H.C., Blaise R.G., Champs M., Pfeiffer B.D.,
RA      Wan X.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA      Abiril J.F., Abpayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA      Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA      Beeson K.Y., Benos P.V., Bertan B.P., Bhandari D., Bolshakov S.,
RA      Borokova D., Botchan M.R., Bouck U., Brokstein P., Brotlier P.,
RA      Burrows K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA      Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA      de Pablos B., Delcher A., Deng Z., Meyers A.D., Dew I., Dietz S.M.,
RA      Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA      Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA      Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasner K.,
RA      Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Guan P., Harris M.,
RA      Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA      Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA      Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA      Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA      Lasok P., Lei Y., Levitsky A.A., Li J., Li Z., Jiang Y., Lin X.,
RA      Liu X., Mattei B., McIninch T.C., McLeod M.P., McPherson D.,
RA      Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA      Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA      Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA      Palazuelo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA      Reinert K., Rillington K., Saunders R.D.C., Scheeler F., Shen H.,
RA      Shue B.C., Sidenkian I., Simpson M., Skupski M.F., Smith T.,
RA      Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA      Svetkey R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA      Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,

```

RA Williams S.M., Woodage T., Worley K.C., Mu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*";
 RL Science 287:2185-2195(2000).
 RN [7]
 RP SEQUENCE OF 959-1078 FROM N.A.
 RC STRAIN=Daekwanryeong;
 RX MEDLINE=85137938; PubMed=2983332;
 RA Madworth S.C., Vincent W.S. III, Blodreau-Wentworth D.;
 RT "A *Drosophila* genomic sequence with homology to human epidermal
 RT growth factor receptor";
 RL Nature 314:178-180(1985).
 RN [8]
 RP SEQUENCE OF 1133-1137, 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
 RP ANALYSIS.
 RX MEDLINE=92038942; PubMed=1936959;
 RA Raz E., Schejter E.D., Shilo B.Z.;
 RT "Interallelic complementation among DER/flb alleles: implications for
 RT the mechanism of signal transduction by receptor-tyrosine kinases";
 RL Genetics 129:191-201(1991).
 RN [9]
 RP REVIEW.
 RX MEDLINE=97248481; PubMed=9094709;
 RA Perrimon N., Perkins L.A.;
 RT "There must be 50 ways to rule the signal: the case of the *Drosophila*
 RT EGF receptor";
 RL Cell 89:13-16(1997).
 CC - FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
 CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
 CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
 CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
 CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
 CC ESTABLISHMENT OF VENTRAL CELL FATES. MAINTENANCE OF AMNIOTEROSA
 CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
 CC CUTICLE.
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC - SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
 CC PROTEIN.
 CC - ALTERNATIVE PRODUCTS: 3 ISOFORMS, TYPE I (SHOWN HERE), TYPE II AND
 CC TYPE III ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC - TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
 CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
 CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
 CC AND THORACIC AND ABDOMINAL GANGLIA.
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF052754; AAC08536.1; -;
 DR EMBL: AF052753; AAC08535.1; JOINED.
 DR EMBL: AF052754; AAC08535.1; -;
 DR EMBL: AF052752; AAC08535.1; JOINED.
 DR EMBL: K03054; AAAS1462.1; -;
 DR EMBL: X03417; AAAS1460.1; -;
 DR EMBL: K03416; AAAS0965.1; -;
 DR EMBL: K03418; AAAS1461.1; -;
 DR EMBL: AF109077; AAD26134.1; -;
 DR EMBL: AF109078; AAD26132.1; -;

DR EMBL: AF109082; AAD26132.1; JOINED.
 DR EMBL: AF109078; AAD26133.1; -;
 DR EMBL: AF109084; AAD26133.1; JOINED.
 DR EMBL: AF109079; AAD26130.1; -;
 DR EMBL: AF109081; AAD26130.1; JOINED.
 DR EMBL: AF109079; AAD26131.1; -;
 DR EMBL: AF109083; AAD26131.1; JOINED.
 DR EMBL: AF109080; AAD26135.1; -;
 DR EMBL: AE003454; AAF46732.1; -;
 DR EMBL: X02293; CA26157.1; -;
 DR EMBL: X78920; CAAS5523.1; -;
 DR EMBL: X78919; CAAS5521.1; -;
 DR EMBL: X78919; CAAS5522.1; -;
 DR PIR: A00640; GORPE.
 DR HSBP; P11562; IFGK.
 DR Flybase; FBgn0003731; Egr.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_Pkinase.
 DR InterPro; IPR002174; Purin-like.
 DR InterPro; IPR001245; Tyr_Pkinase.
 DR Pfam; PF00069; Pkinase_1.
 DR Pfam; PF00757; Purin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_Pkinase; 1.
 DR SMART; SMO0261; FU_7.
 DR SMART; SMO0219; TykC; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR Trasnembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
 DR Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
 DR Developmental protein.
 DR SIGNAL; 30.
 DR CHAIN; 31; 1426; POTENTIAL.
 DR DOMAIN; 31; 868; EPIDERMAL GROWTH FACTOR RECEPTOR.
 DR TRANSSEM; 869; 889; EXTRACELLULAR (POTENTIAL).
 DR DOMAIN; 890; 1426; POTENTIAL.
 DR DOMAIN; 938; 1198; CYTOPLASMIC (POTENTIAL).
 DR NP_BIND; 944; 952; PROTEIN KINASE.
 DR BINDING; 971; 971; ATP (BY SIMILARITY).
 DR ACT_SITE; 1063; 1063; ATP (BY SIMILARITY).
 DR MOD_RES; 902; 902; BY SIMILARITY.
 DR PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 Query Match 28.2%; Score 1935; DB 1; Length 1426;
 Best Local Similarity 32.4%; Pred. No. 4,3e-96;
 Matches 464; Conservative 187; Mismatches 435; Indels 344; Gaps 42
 QY 24 QVCTGDMKRLPASPETHIDMLRLHYGGVQGNLELTYLPT-NASLSPLODIOEVQV 82
 DB 100 KICGTGSRSLSPVSNKHEHRYNRDRYCTGYVDGMLKTLWLPENYLDLSFLDNIREVTV 159
 QY 83 YVIAHQVQVQLPQRLRIYRGQLF-----ENVYLAFLDNDPLNNTPTPGASPGSL 137
 DB 160 YILSHVDKRVKVPFKQILIRGTLTSLSVSEKVALFV-----TYSKM 203
 QY 138 RELQRLSLLEILKGVLYLQRPOLCYODITLWKDIFKKNQALTLIDTNRACHPSCP 197
 DB 204 YLEIPLRLDVLNQGQGFHNHYVLGMRITQSEIYNSNTDAVYNYDFAPRECPKCHE 263
 QY 198 MCKGSRWQSSSSDQSLRTVYAGGCA--RCKGPIPTDCCHGQCAAGTGPSPHSDCLAC 255
 DB 264 SCTHG-CWSEGPKNCKKFSFLTSSPCACAGRCVGPAPRECHLFCAGGCTGPTORDCIAC 322
 QY 256 LHFHSGICELCPALVTYNTDFFSEMPNPEGRYFGASCYTRACPYNLTSTVGSCITLVC 315
 DB 323 KNFDEAVSKEBCPRKRYNPTTYLVETNPEKRYAGATVACEP-GHLLRNGACVASC 381
 QY 316 PLNQEVTAEADGTQREKSKPCARVCYGLGMNNFTVSWILRVKVASHLEEPAGCKK 375
 DB 382 PDKMKKGGE-----CVPCNGCPKTCPG-----VTVLAAGNIDSPRNTCTV 422
 QY 376 IFGSLAFLPESFDG--DPSAN--TAPQYIKANSK-----FIGITELTGLVLYISAWEDSLPD 427

```

Db 423 IDGNRIIDQTFSSGFQDYVANYTWGPRYIPLDPERREVSTVAIEIGYNIIEGTHPQFN 482
Qy 428 LSVFQNLQVIGRILHNGAY-SLTLOGISWGLRSRLRELSGALIHNTLCTVHV 486
Db 483 LSYRNLETTHGRQLMSMFALAIVKSYLSLEMRLLKQISSGVVIOHNRDLCTVSMI 542
Qy 487 PWDQLFRNPQALJHTANRPEDECVGEGLAHQULCARHMGCGPPOCVVCSQFLGQSC 546
Db 543 RMPALQKEPEKQVWVWNEMLRADLEKNGTICSDQCNEDGCMGAGTQCLTKNFNNGTC 602
Qy 547 VEEGRVLOGLEPREVYNAHCLPGHPECOPONGSVTCGPPADOCVACAHYKDPFCVARC 606
Db 603 IADCGYISNAVK--FDNRCKICHPBCR-----TCMGAGADHCQCEVHRDQGHCVSEC 654
Qy 607 P-----SGVK-----PDL 614
Db 655 PKKYNDRGVCRECHANCDCGTGKDTIGAGCTTCNLAINNDATVKKCLDKDKCPD- 713
Qy 615 SYMPIWKF--PDEEGACQP-----CPI-----NCTH----- 638
Db 714 GY--FMEXVHPQEGSKPLAGRAVCRKCHPLCBLCTNYGHEQVCSKTHYKRRQCE 771
Qy 639 -----SC-----VLDLKG----- 647
Db 772 ECPADHTDEQRCFQHPBECNCGTGPADCKSCNFKLPANETGPVYNSMTNCTS 831
Qy 648 -CPAEOR-----ASPLTS-----IVSAVGIILVVVGVVFGI 679
Db 832 KCPLEMRHVNYQYTAIGPYCAASPPRSKITANLDVNMIIIGAVLVPTICLCV--T 889
Qy 680 LIKRQCKIRKYT--MRLLQETLVEPLTPSGAMPNQAOMRLKETLRKVKVLSGAF 737
Db 890 YICQKQAKKETVMKMAISGCDSPKPSNIGALCTLRVYKAEARKGCVLGMGAF 949
Qy 738 GTVYKGIIMPGEVNVKIPVAIKVLRNTSPKANKEIDEAHYVAGVSPYSRLIGICLT 797
Db 950 GRVYKGVWPEGEVNVKIPVAIKELIKSTAESSEFIREVYINASEHVNILKTLAVCMS 1009
Qy 798 STVOLVTQLMFYGCLDVRNENRGLGSDLLMWCMQIAKMSYLEDVRLVHRDLAARV 857
Db 1010 SOMMLITQMLPGLCLDLYVNNRDKISKALLNWSQIAKMSYLEBEKRLVHRDLAARV 1069
Qy 858 LVKSPNHVKITDEGLARLDIDEHYADGKVPKIMMALESLRRFTHQSGVWSYCT 917
Db 1070 LVQTPSLVKITIDFGLAKLSSDSNEYPAAGGKMPKMLAEICIRNVEFTSKSDVMAFGVT 1129
Qy 918 VVELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITIDVYIMVKMMIDECRPRFE 977
Db 1130 IMELLTFGQRPHEMIPAKDIPDLIEVGLKLEPEICSLDITCLLSCHMLDAMRPTFKQ 1189
Qy 978 LVSEFSKARDPQRFVVIQNEDELG--PASPLDSTFTRSLLEDG--DMGDLVDAEYLV 1032
Db 1190 LTVFAEFARDPGYVLIIPGDKFTRJPA-----YTSODEKLIRKLAPTTDGSALRK 1242
Qy 1033 QCGFFCDDPAPAGAGMVRHHRSSSTRSGGGLTLGLEPSEEPAP-----RSLAPSRG 1086
Db 1243 PDDYLOPKAARGP-----HRTDCT-----DMRPLNXYCDDPSKXNS 1281
Qy 1087 AGSDVFDG---DICMGAKGLQSLPTHDEPFLQRYSDPVPVLPSTSDGYVAALTCSPQ 1143
Db 1282 TGDDEKSSAREVGVGLR-----LDLPVDEDDYLLMP--TCQPPQ 1319
Qy 1144 EYVNPDPVRPQPSRPGPLPAARPAQATLERAKTLSPKNGVGVVAFPGAGVENPEVL 1203
Db 1320 NNNNNNN-----NPNONMAAIVGAAGVY-----DLIGVPVSDVDEPIL 1358
Qy 1204 ----TPGGGAAPQPH-----PPAFSP-AEDNLVYWD 1230
Db 1359 LNAQTIGVGESPIPTQTIGIPVWGPGQTMVVKVPMGSEPTSDHEVYND 1408

```

RESULT 12

```

ERBB ALV
AC _ERBB ALV STANDARD; PRT; 634 AA.
ID P00534
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUN-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN
OS Avian leukosis virus.
OC Viruses; Retroviruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nielsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Crichtenden L.B., Raines M.A., King H.-J.,
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor."
RL Cell 41:719-726(1985).
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -I- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -I- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboratio
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on it
CC use by non-profit institutions as long as its content is in no wa
CC modified and this statement is not removed. Usage by and for commercia
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M10066; AAA48763.1; ALT_INIT.
DR PIR; A00643; TVCHLV.
DR PIR; B00643; TVFVLV.
DR HSSP; P11362; IFKG.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00069; Pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Euk_Pkinase; 1.
DR SMART; SMO0219; TyrcK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR TRANSFERASE; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
SQ
SEQUENCE 634 AA; 70891 MW; E705E33A0B01FCC CRC64;

Query Match 25.5%; Score 1749.5; DB 1; Length 634;
Best Local Similarity 52.3%; Pred. No. 1.4e-86;
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17

Qy 593 CAHYKDPFCVARCPSGVKPDLSYMPIWKFPEDEGACQCPINCTHSCVLDLDDKCCPAEQ 652
Db 3 CAHFIDGPHCVKACPAAGVGENDTL-VWKYADANAVCQLCHNCTRGCKGPELBECP- 58
Qy 653 RASPLSTISAVV-GILLVWLVGVNGLIKRQCKIRKYTKRRLQELVELVEPLTPSGA 711
Db 59 NSKRTSIAAGVGVGLLCLVVGGLGVLRR-HIVRRTLRLLQERELEVEPLTPSGE 117
Qy 712 MPNQAOMRLKETLRKVKVLSGAGFVYKGIWIPDGENVAKIPVAIKVLRNTSPKANX 771

```

```

Db 118 APNOAHLRIKETEPKVKVLSGAFGYKGLMIPEGEKXIPAIKELREAFSPANK 177
QY 772 EILDEAYVMAGVSPVSRLLGICLTSTVQVLTQMLPFGCLLDHVRNKGRLSGODLLNW 831
Db 178 EILDEAYVMASVDNPHVCRLLGICLTSTVQVLTQMLPFGCLLDYIREHKDINGQVLLNW 237
QY 832 CMQIAGKMSYLEEDVRLVHRDLAARVNLVKSPPHVKITDFGLARLLDIDETFEHADGGKVP 891
Db 238 CVQIAGKMWYLEERLVRDLAARVNLVKTPOHVKITDFGLAKLLGADKEXYHAEGRKVP 297
QY 892 IKMMALESILRRRFLTHQSDVMSYGVTVWELMTFGAKPPDGIIPAREIPDLLEKGERLPQPP 951
Db 298 IKMMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPPDGIIPASEISSVLEKGERLPQPP 357
QY 952 ICTIDVYMIWVKCMNIDSECRPRELVSFSPMARDPQRFVVIQ-NEDIGPASPDLSTF 1010
Db 358 ICTIDVYMIWVKCMNIDADSRKRFELIAESKMARDPRIYLVQGBRMLPSPDTSKF 417
QY 1011 YRSLEDDDMGDLVDAEYLVPOGFFCPDPAPAGAGMWHRRSSSTRSGGDLTLGLE 1070
Db 418 YRLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
QY 1071 PSEEARPSPL-----APSEGASVDFDGLGMAKGLQSLPTHDSPLQRYSEDPVP 1125
Db 450 -----SRTPLLSSAISATSNNSATCID-----RNGGHPVREDSFVQRYSSDPTGN 495
QY 1126 LPSET--DGVAFLTCSPOPEVYNOPDVRPOPSPRESGPLPAAPAGATLERAKTLSFGK 1183
Db 496 FLEESIDGFL-----PAPEYVNG--LMPKKPS-----TWMQ 526
QY 1184 NGVAKDVF-----AEGAVENPEYITPQGGAPPPHPPAPSPAFDNLVY 1228
Db 527 NOIYNNISLTAISKLPMSRYONSHTVADNPEYL-----NTNOSPAAKTFFESSPY 578
QY 1229 WDQ-----DPPE-----RCAPSTFGKTPAENPEYLGIDVP 1260
Db 579 WTQSGNHQINLNDNPYQODFLPNTKRNGLKVPRAENPEYLRVAAP 625

RESULT 13
ERBB_AVIER STANDARD; PRT; 604 AA.
AC P00335;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RA MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbb gene of avian erythroblastosis virus is a member of the src
RT gene family."
RL Cell 35:71-78(1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RA MEDLINE=8423957; PubMed=6328658;
RA Debutre B., Henry C., Benasassa M., Biserre G., Claverie J.-M.,
RA Saulie S., Martin P., Stehelin D.;
RT "Sequencing the erba gene of avian erythroblastosis virus reveals a
RT new type of oncogene."
RL Science 224:1456-1459(1984).
CC -i- Catalytic Activity: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -i- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.

```

```

CC -i- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -i- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaborati
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on it
CC use by non-profit institutions as long as its content is in no w
CC modified and this statement is not removed. Usage by and for commerci
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1; -.
DR PIR; A00644; TVYUH.
DR HSSP; P13562; IREGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP. 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR. 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM. 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76EBD0D6745D609 CRC64;

Query Match 24.9%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 4e-84;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 14
QY 593 CAHYKDPFCVACRCPGVPPLDSYPMKFPDEGACQPCPINCTHSCVDLDDKCPAQ 652
Db 3 CAHFTDGHCVKACRAGVLGENTLT-VKXADAIVACQLCHPNTCRCKPGGLGECF-- 58
QY 653 PASPLTGSVSAVY-GILIVVLGVVFGILIKRQOKIRKTYMRRLLOETELVEPLTPSGA 711
Db 59 NGSKTPSTAGAVGGLLCTLVVVGIGILYLRK-HYRKETLRLLQERLVEPLTPSGE 117
QY 712 MPNOQMRILKETELRKVKVLGSGAFGVYKGIWIPGGEVYKIPVAIKVLRNTSPANK 771
Db 118 APNOAHLRIKETEPKVKVLSGAFGYKGLMIPEGEKXIPAIKELREAFSPANK 177
QY 772 EILDEAYVMAGVSPVSRLLGICLTSTVQVLTQMLPFGCLLDHVRNKGRLSGODLLNW 831
Db 178 EILDEAYVMASVDNPHVCRLLGICLTSTVQVLTQMLPFGCLLDYIREHKDINGQVLLNW 237
QY 832 CMQIAGKMSYLEEDVRLVHRDLAARVNLVKSPPHVKITDFGLARLLDIDETFEHADGGKVP 891
Db 238 CVQIAGKMWYLEERLVRDLAARVNLVKTPOHVKITDFGLAKLLGADKEXYHAEGRKVP 297
QY 892 IKMMALESILRRRFLTHQSDVMSYGVTVWELMTFGAKPPDGIIPAREIPDLLEKGERLPQPP 951
Db 298 IKMMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPPDGIIPASEISSVLEKGERLPQPP 357
QY 952 ICTIDVYMIWVKCMNIDSECRPRELVSFSPMARDPQRFVVIQ-NEDIGPASPDLSTF 1010
Db 358 ICTIDVYMIWVKCMNIDADSRKRFELIAESKMARDPRIYLVQGBRMLPSPDTSKF 417
QY 1011 YRSLEDDDMGDLVDAEYLVPOGFFCPDPAPAGAGMWHRRSSSTRSGGDLTLGLE 1070
Db 418 YRLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
QY 1071 PSEEARPSPL-----APSEGASVDFDGLGMAKGLQSLPTHDSPLQRYSEDPVP 1125

```



```

Db      450  -----SRTLLSLTSLSATSNNSATNCID-----RMGQGHFVREDSFYQRSSDPTGN 495
Oy      1126 LBSEET--DGVAVALTCSPPPEYVNQPDVRPQPSPSPREGPLPAARPGACATLERAKTISPEK 1183
Db      496 FLEES:LDDGL-----PAPEYVNO--LMPKKPFSTAM-----524
Oy      1184 NGVAVKDFAF-----GGAVENPEYLTPFGGAAPQHPPEAFSPARD 1224
Db      525 --VONQIVNYSITLATSKLPMOSRYQNOSHSTRAYVNPELL-----NNQSPLAKTYVE 574
Oy      1225 NLYVMQDPPEPEGAPSPSTFKGPIFTENEKY 1254
Db      575 SSPWYIQSGNHQ-----INDNDY 594

RESULT 14
ERBB_AVIEU
ID_ERBB_AVIEU STANDARD; PRT: 540 AA.
AC P11273;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
CN v-ERBB.
OC Avian erythroblastosis virus (strain ts167).
OS Viruses; Retrocid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=103898;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87064459; PubMed=2878934;
RA Choi O.R., Trainor C., Graf T., Beug H., Engel U.D.;
RT "A single amino acid substitution in v-erbB confers a thermolabile phenotype to ts167 avian erythroblastosis virus-transformed erythroid cells."
RL Mol. Cell. Biol. 6:1751-1759 (1986).
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.
CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC CC
DR EMBL; M13179; AAA42401.1; -.
DR PIR; A25231; TVVEBH.
DR HSPD; P11362; IFGR.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD0000001; Euk.pkinase; 1.
DR SMART; SMO0219; TYRKc1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP_1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR_1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM_1.
KW Transerase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).
SQ SEQUENCE 540 AA; 60412 MW; 5B53297AA06B65D CRC64;

Query Match 23.8%; Score 1630; DB 1; Length 540;
Best Local Similarity 55.1%; Pred. No. 2.8e-80;
Matches 341; Conservative 69; Mismatches 119; Indels 90; Gaps 14;

```

Db	CAHRTDGHCVKACPGAVGLGENDTL-VWKYADANNAVQDCHPNCTRGGKGGLEGCP----	58
Qy	653 RASPLTIVSAVY-GILLVYVIGVVFGLIKRROCKIRKYMRLLOETLVEPLVTSGA	711
Db	59 NGSTPTPIAGVGVGGLCLVWVGLG:IGLTRRR-HIVKRTLRLRLORELVETPSGE	11
Qy	712 MPNOAKRIKETELKRVKVLGSGAAGTYKQIM:PDGANKYIPAIYVIRENTSPANK	772
Db	118 APNOAHLRIKETELKRVKVLGSGAAGTYKQIM:PESEKTIYPAIKELREATSPANK	177
Qy	772 EILDEAYVMAGVSPYVSRLLGICLTSTVQVLVTQMLPYGCLLDHRENRGLGSDLLNW	831
Db	178 EILDEAYVMASVNPVPCVRLGICLTSTVQVLITQLMPCGLLDVIREHKDNIQGYLLNW	237
Qy	832 CMOIAKMSYLEVRLVHRD:LAARNVLYKSPHNVITDTPGLARLLDIDETVHADGGKVP	891
Db	238 CVQIAKMSNYLREHRVHRD:LAARNVLYKTPQHVITDTPGLAKQGADEKEYHAEGGKVP	297
Qy	892 IKMAALESILRRFRTHQSDVMSYGYVTVELMTPGAKPYDGI:PAEIPDLLEKGERLPDP	951
Db	298 IKMAALESILHRIYTHQSDVMSYGYVTVELMTPGSKPYDGI:PAEISISVLEKGRRLQPP	357
Qy	952 ICTIDVYIMVWKCMWMDISCRPFELVSEFSKMARDPQRFVVIQ-NEDUGPASPLDSTF	1011
Db	358 ICTIDVYIMVWKCMWMDADSRKRFELIAEFSKMARDPRLVYIQGERMILPSPDTSKF	417
Qy	1011 YRSILLEDMDMDGLVDAEEYLVVQOQGFPCPDPAFGAGVHHRHSSSTRSGGDLTGLLE	1071
Db	418 YRTLMBEDMDMDGLVDADEYLVVHQGF-----NSPST-----	445
Qy	1071 PSEBEARPSPL----APSEGAQSVFPGDGLMGAKGLQLPHDPSPLQRYSEDPVTP	1121
Db	450 -----SRPTLLSLSLATSNNSATCIDRNQ-----H-----	476
Qy	1126 LPSETDGYVAFPLCSQPEYVQNPDPQPSPREGPI:PAAPAPGAT-LEBAKTLSPGX	1186
Db	477 -PREDGFL-----PAPRYVQ--LMPKISTAMVQNIQITNY:SLTAISKLPIDSKYQN	527
Qy	1185 GVVYKDVAFAGGAVENEPEYL 1203	
Db	528 -----SHSTAVDNEPEYL 539	
RESULT 15		
EGFR_CHICK		
ID	EGFR_CHICK STANDARD; PRT; 703 AA.	
AC	P13387;	
DT	01-JAN-1990 (Rel. 13, Created)	
DT	01-JAN-1990 (Rel. 13, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER	
DE	(Fragment).	
GN	EGFR.	
OS	Gallus gallus (Chicken).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Aukosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;	
OC	Gallus.	
OX	NCBI_TaxID=9031;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=88261272; PubMed=3260329;	
RA	Lak I., Johnson A., Howk R., Sapp J., Bellot F., Winkler M.,	
RA	Ullrich A., Vannestrom B., Schlesinger J., Givol D.:	
RT	"Chicken epidermal growth factor (EGF) receptor: cDNA cloning,	
RT	expression in mouse cells, and differential binding of EGF and	
RT	transforming growth factor alpha."	
RL	Mol. Cell. Biol. 8:1970-1978(1988).	
CC	-1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,	
CC	AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND	
CC	VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).	
CC	-1- CATALYTIC ACTIVITY: ATP -> a protein tyrosine = ADP + protein	
CC	tyrosine phosphate.	
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.	

CC	-1-	MISCELLANEOUS:	Binding of EGF to the receptor leads to dimerization, internalization of the EGF-receptor complex, induction of the tyrosine kinase activity, stimulation of cell DNA synthesis, and cell proliferation.
CC	-1-	SIMILARITY:	BELONGS TO THE EGF RECEPTOR FAMILY.
CC	CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/stb-stb.ch).	
CC	CC	or send an email to license@isb-sib.ch .	
DR	EMBL:	M20386; AAA48760.1;	-
DR	InterPro:	IPR000494;	EGF_R_L_domain.
DR	InterPro:	IPR000719;	Euk_pkinase.
DR	InterPro:	IPR002174;	Furin_kinase.
DR	Pfam:	PF00757;	Furin-like; 1.
DR	Pfam:	PF01030;	Recep_L_domain; 2.
DR	SMART:	SM00261;	FU; 4.
DR	PROSITE:	PS00107;	PROTEIN_KINASE_ATP_PARTIAL.
DR	PROSITE:	PS00109;	PROTEIN_KINASE_TYR_PARTIAL.
DR	PROSITE:	PS50111;	PROTEIN_KINASE_DOM_PARTIAL.
KM	Transmembrane;	Glycoprotein; Receptor; Signal; Transferase;	
KW	Tyrosine-protein kinase;	ATP-binding; Phosphorylation.	
FT	SIGNAL	1	30
FT	CHAIN	31	>703
FT	DOMAIN	31	654
FT	TRANSEM	655	667
FT	DOWNLN	668	>703
FT	DISULFID	197	206
FT	DISULFID	201	214
FT	DISULFID	222	230
FT	DISULFID	226	238
FT	DISULFID	239	247
FT	DISULFID	243	255
FT	DISULFID	258	267
FT	DISULFID	271	298
FT	DISULFID	302	314
FT	DISULFID	318	333
FT	DISULFID	336	340
FT	DISULFID	515	522
FT	DISULFID	517	530
FT	DISULFID	533	542
FT	DISULFID	546	562
FT	DISULFID	565	581
FT	DISULFID	569	589
FT	DISULFID	592	601
FT	DISULFID	605	627
FT	DISULFID	630	638
FT	DISULFID	634	646
FT	CARBOHYD	134	134
FT	CARBOHYD	190	190
FT	CARBOHYD	200	200
FT	CARBOHYD	359	359
FT	CARBOHYD	368	368
FT	CARBOHYD	420	420
FT	CARBOHYD	573	573
FT	CARBOHYD	578	578
FT	CARBOHYD	613	613
FT	CARBOHYD	633	633
FT	CARBOHYD	648	648
FT	NON_TER	703	703
QO	SEQUENCE	703 AA;	77427 MW; AFD2DE1B735A690 CRC64;

```

Query Match      22.7%; Score 1558; DB 1; Length 703;
Best Local Similarity 43.3%; Pred. No. 2.7e-76;
Matches 309; Conservative 113; Mismatches 257; Indels 34; Gaps 13;

      6 RMGLIALPFGAA-----SYQVCTGSDMYLRAPAFETNIDMLNHYGGCGVQGNIE 61
      * : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

[illegible]

Search completed: July 22, 2003, 08:46:20
Job time : 21.2793 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.5887 Seconds

(without alignments)
5347.444 Million cell updates/sec

Title: SEQ-369-383-14

Perfect score: 6853

Sequence: 1 MELALCRMGLLALLPFGA.....TKGTPTAENPYGLDVPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6154	89.8	1259	6	018735
2	3102	45.3	1209	1	09GX70
3	3073	44.8	1210	11	09EP98
4	2688	39.2	1165	13	09YH40
5	2660.5	38.8	1137	13	09G6F6
6	2280	33.3	1338	13	P79754
7	2011.5	29.4	1433	5	09B1H9
8	1871	27.3	419	4	09UX79
9	1739	25.4	367	11	08R2X1
10	1720	25.1	729	15	08G712
11	1718	25.1	567	15	08G714
12	1697.5	24.8	412	4	08WV0
13	1654	24.1	962	15	084895
14	1645	24.0	545	15	085468
15	1464.5	21.4	655	11	09WVF5
16	1448.5	21.1	643	11	09ERV6

17	1249	18.2	1193	5	09Y1X8	09Y1X8 ephradytia f
18	1188.5	17.3	1368	5	023821	023821 caenorhabdi
19	1179	17.2	1717	5	026566	026566 schistosoma
20	1089	15.9	527	13	090836	090836 gallus gall
21	959.5	14.0	478	11	09SEB0	09SEB0 rattus norv
22	906	13.2	165	4	014256	014256 homo sapien
23	905.5	13.2	599	13	09PSH2	09PSH2 gallus gall
24	887	12.9	176	11	0923V5	0923V5 rattus norv
25	806.5	11.8	346	13	P11776	P11776 xiphophorus
26	778	11.4	435	5	08SZM1	08SZM1 drosophila
27	754.5	11.0	311	13	09P162	09P162 xiphophorus
28	739.5	10.8	1362	13	09PVZ4	09PVZ4 xenopus lae
29	734	10.7	331	4	09BUD7	09BUD7 homo sapien
30	725	10.6	1671	5	09NV5	09NV5 biophalari
31	723	10.6	149	6	09BC66	09BC66 oryctolagus
32	687	10.0	1418	13	093457	093457 scophthalmu
33	683.5	10.0	1368	13	08UM85	08UM85 paralicthy
34	666.5	9.7	1369	13	08UM86	08UM86 paralicthy
35	665	9.7	1422	13	08UM84	08UM84 paralicthy
36	663.5	9.7	1472	5	09U5A8	09U5A8 bombyx mori
37	657	9.6	1358	13	073798	073798 xenopus lae
38	643.5	9.4	1418	13	08UM83	08UM83 paralicthy
39	628	9.2	1245	13	09YGH8	09YGH8 scophthalmu
40	624	9.1	1371	11	09QVW4	09QVW4 rattus sp.
41	618.5	9.0	2144	5	09VD94	09VD94 drosophila
42	550.5	8.6	967	11	09QWR2	09QWR2 mus musculu
43	587.5	8.6	987	11	091YMO	091YMO mus musculu
44	587.5	8.6	1036	4	007912	007912 homo sapien
45	587.5	8.6	1055	11	054967	054967 mus musculu

ALIGNMENTS

RESULT 1
ID 018735 PRELIMINARY; PRT; 1259 AA.
AC 018735;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Erbb-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
CX NCBI_TaxId=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008451; BAA23127.1; -;
DR HSBP; P11362; IFGK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR L domain.
DR InterPro; IPR000719; Euk. kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr-kinase.
DR InterPro; IPR04019; YLP-motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR Prodom; PD000001; Euk. kinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TYKC; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D9C4ACD45 CRC64;

Query Match	89.8%; Score 6154; DB 6; Length 1259;
Best Local Similarity	89.8%; Pred. No. 0;
Matches 1138; Conservative	48; Mismatches 67; Indels 14; Gaps 4;
QY	1 MELALCRWGLLALLPFGAASSTVCTGTDMLRLPASPEHLLMLRLHYQGCQVVGNI 60
DB	1 MELAMCRWGLLALLPFGAAGTVCCTGTDMLRLPASPEHLLMLRLHYQGCQVVGNI 60
QY	61 ELTYLPTNASTSLFDIOIEVQGVYLIANQVRYOYPLRLRIVRGTOLEFENYALAVLDNG 120
DB	61 ELTYLPANASTSLFDIOIEVQGVYLIANQVRYOYPLRLRIVRGTOLEFENYALAVLDNG 120
QY	121 DPLNNTVTGTASPGGLRELOLRSLTEILKGVIIQGNPOLCYQDTLLMKDIFHKNOQA 180
DB	121 DPLGGIIPAGMAOQGLRELOLRSLTEILKGVIIQGRSPQLCHQDITLLMKDVFHKNOQA 180
QY	181 LTLIDTNRBRACHPSPMCKGRMCGSESSCCSLTTCVAGGARCKGLPTDCCHEQC 240
DB	181 LTLIDTNRFSACPSPSPACKDAHCKWASSGDCQSLITVCAAGGARCKGPPTDCCHEQC 240
QY	241 AAGCTGPRGSDCLALFHNHSGICELHCPALVTYNTDTPESMNPREGRYTFGASCVTACP 300
DB	241 AAGCTGPRGSDCLALFHNHSGICELHCPALVTYNTDTPESMNPREGRYTFGASCVTSCP 300
QY	301 YNYLSTDVGSCITVCPHNCVTAEDGTQRCCKSKCAVCYGLMNNFTYSFWLR-V-359
DB	301 YNYLSTDVGSCITVCPHNCVTAEDGTQRCCKSKCAVCYGLMNNFTYSFWLR-V-359
QY	360 PKVASHLLEFAGCKITFGSLAFIPESFDDGPASNTAPQYIKANSKFGITETGLYIS 419
DB	360 PKVASHLLEFAGCKITFGSLAFIPESFDDGPASNTAPQYIKANSKFGITETGLYIS 419
QY	420 AMPDPLPLSYFOULQYIRGRILHNGAYSLTLOGLISWGLRSLBELSGCALIHHNTH 479
DB	420 AMPDPLPLSYFOULQYIRGRILHNGAYSLTLOGLISWGLRSLBELSGCALIHHNTH 479
QY	480 LCFVHTVPMDOLEFNPQALHTANRDEDEGEGALACHOLCARGCMGPGPQCNCQ 539
DB	480 LCFVHTVPMDOLEFNPQALHTANRDEDEGEGALACHOLCARGCMGPGPQCNCQ 539
QY	540 FLRQGECEVCEGVYQGLPREYVNAHCLPCHPECPQNSVTCFGEADQCVACAHYKDP 599
DB	540 FLRQGECEVCEGVYQGLPREYVNAHCLPCHPECPQNSVTCFGEADQCVACAHYKDP 599
QY	600 PFCVACPSGVKPDLSMPIMKPEDEGACOPCINTHSCVULDKGCPADGRASPLTS 659
DB	600 PFCVACPSGVKPDLSMPIMKPEDEGACOPCINTHSCVULDKGCPADGRASPLTS 659
QY	653 IIAVVGILLAVVGLVIGILIKRRQKIRKYMRLLOTELEVEPLTSGAMPQAKWR 712
DB	653 IIAVVGILLAVVGLVIGILIKRRQKIRKYMRLLOTELEVEPLTSGAMPQAKWR 712
QY	720 ILKETELRKVYVLSGSAFGTVYKGIWIPDENVKIPAIKVALENTSPANKIIDEAVY 779
DB	720 ILKETELRKVYVLSGSAFGTVYKGIWIPDENVKIPAIKVALENTSPANKIIDEAVY 779
QY	773 MAGVGSFYVSRLLGICLTSTVQVLTQMPRGCLLDVREHRLGSGODLIMNCVQIAKAM 832
DB	773 MAGVGSFYVSRLLGICLTSTVQVLTQMPRGCLLDVREHRLGSGODLIMNCVQIAKAM 832
QY	840 SYLEDVRLVHRDLAARVYLYKSPHVKITDFGLARLLDDETEYHAGDGKVPDKMALES 899
DB	840 SYLEDVRLVHRDLAARVYLYKSPHVKITDFGLARLLDDETEYHAGDGKVPDKMALES 899
QY	900 ILRRRFTHQSVDVSYGTVWELMTFGAKPYDGIIPAREIDLLLEKGRILPOPICTIDVYM 959
DB	900 ILRRRFTHQSVDVSYGTVWELMTFGAKPYDGIIPAREIDLLLEKGRILPOPICTIDVYM 959
QY	960 IMYCKMWIDSECPRRFELVSESKMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDD 1019
DB	960 IMYCKMWIDSECPRRFELVSESKMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDD 1019
QY	1020 MGDVDAEYLVVQGGFFCPDPAFGAGMWHHRSSSTRSGGDLTLGLEPSEBEAPRS 1079

DB	1013 MGDVDAEYLVVQGGFFCPDPAFGAGTARHRNRSSSTRSGGDLTLGLEPSEBEAPRS 107
QY	1080 PLAPSEAGSDVDFDGLGMGAAXKGLSLPTHDSPLOQRYSEDETPVLPSETDGVAPLTC 113
DB	1073 PLAPSEAGSDVDFDGLGMGAAXKGLSLPTHDSPLOQRYSEDETPVLPSETDGVAPLTC 113
QY	1140 SPOPEYVNPQVPRQPPSPREGPLPAARPAATLTER-----AKTISPKKGVYKDVPAFG 115
DB	1133 SPOPEYVNPQVPRQPPSPREGPLPAARPAATLTERPKTSLSPGKGVYKDVPAFG 115
QY	1195 GAVENPEYLTPOGGAAPQPPAPFAPFNLYYMDODPERGAPSTFGTPTAENPEY 125
DB	1193 SAVENPEYLTARGAAPQPPAPFAPFNLYYMDODPERGAPSTFGTPTAENPEY 125
QY	1255 LGLDVVPV 1261
DB	1253 LGLDVVPV 1259

RESULT 2

090X70

ID 090X70 PRELIMINARY; PRT; 1209 AA.

AC 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Epidermal growth factor receptor.

OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RX MEDLINE=9025888; PubMed=2342466;
 RA Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,
 RA Earp H.S.;
 RT "A truncated, secreted form of the epidermal growth factor receptor is
 RT encoded by an alternatively spliced transcript in normal rat tissue.";
 RL Mol. Cell. Biol. 10:2973-2982(1990).
 RN [2]

RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RA Petch L.A.;
 RL Submitted (NOV-1991) to the EMBL/Genbank/DBJ databases.

DR InterPro: IPR000494; EGFR_L domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF01030; Recep_L domain; 2.
 DR PRINTS: PR00109; TYRKINASE.
 DR PRODOM: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; FU; 3.
 DR SMART: SM00219; TYKc; 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE: PS00101; PROTEIN KINASE DOM; 1.
 DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.
 KM ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 1209 AA, 134691 MW, 365877cC18773 CRG64;

Query Match 45.3%; Score 3102; DB 11; Length 1209;
 Best Local Similarity 49.7%; Pred. No. 1,3e-225;

```

Matches 639; Conservative 166; Mismatches 356; Indels 124; Gaps 29;

QY 3 LAALCRWGLLALPPGA-ASTOVCTGDMKLRLPASPETHLDMLRLHYGCGVVGQNLIE 61
Db 15 LRAALCAAG-----GALBEKKVCGGTSNRLTQLGTFEDHFLSLQMFNNCEVVGANLE 66
QY 62 LTYLPTNASLSFLQDIOEVQGVLIANNOVQLRIYRGTOLEEDYVALAVLDNGD 121
Db 67 ITTVQRNYDLSLFKIQEAGVYLIALYTERIFLENQIIRGNALYENTYALAVLSN-- 124
QY 122 PLNNTPTVGSFGLRELQRLSLTEILKGVLLQRPOLCYQDTIMWQIFKKNQAL 181
Db 125 -----YGNKTKGLRELPMNNLOEILIGAVRSPNNPILCNMETIQWRDVI--QDFLSN 175
QY 182 TLIDTNRS-RACHPCSPMCKGSRGCGESSDDQSLTRTVACGCA-RCKPLPTDCHEQ 219
Db 176 MSMQVQRHLTGCPKCDPSCPNGSCWGGEENCQGLTKITICAGQCSRRCRSRSDCHNQ 225
QY 240 CAAGCTGPKHSDCLALHFNHSGICEHCPALVTYNTDTESMNPFGRTTFGASCYTAC 299
Db 236 CAAGCTGPRESDCLVCHRFDEATCKTCPEPLMYNPTTYQMDVNPBGKYSFGATCVKXC 295
QY 300 PYNLTSDVGSCTLVCPLNQOEVTAEDGTORCEKSPCARVCYGLGM--FNNFTVSPML 357
Db 296 PRNYVYTDHSGCVARACGPDYEV-EDGVSKCKKCDGPRCKVCGIGIGFKD-TLS--- 350
QY 358 RVPKVASHLEEPGCKKIFGSLAFLESFGDPASNTADQVIKANSKFIIGITELGYLY 417
Db 351 ----INATNIGHFKYCTAISGDLHLPLVAFKGDSEFTTPELPRELEILKTXVEITGFL 406
QY 418 ISAMPDSLPLDSVFQNLQVIRGRILHNGAVSLTLOGIGIMLGIRSLREGSLALIHNN 477
Db 407 IQAMPENMTDLHAEENLEIIRGTRKQHQGSLAVAGNLTSLGRSKIESLSDGVITISGN 466
QY 478 THLCFHTVPMQDLFRNPHQALLTANRPEDECVEGELACHOLCARHCWGPPTQVNC 537
Db 467 RNLQYANTINMKKLFQPNCKTKIMNNRAEKDKATNVHCNPICSSGCGGPEPTQVGC 526
QY 538 SQFLRGCEVECEVLOGLPREYVNAHCHLPCHEQCQPGQSGVTCGPEPEADQVCAHRYK 597
Db 527 QNVSRGECVCKVILRGEPREFENSECICQCHECIPQMTNITCGRGPDNCKIKAHAV 586
QY 598 DPFCVACPSGVAPDLSYMPIMKFPDEGACQCPICNTHSCVVDLDDKCPAEOBASP- 656
Db 587 DGHCVKTCPSGINGENNTL-VMKFADANNVCHLCHANCTVYGAGGILKGC--QDEBGRK 643
QY 657 LTIYSAVVGLLVVLCVVFGL-LIKRQCKIRKTYMRPLLOETELVEPLTPSGAMPQ 715
Db 644 IPSIATGIVGLRIYV-VALGIGLFMRKROLVAKRTLRLLQERELVEPLTPSGEAPNQ 702
QY 716 AQMRILKETELRKVYLSGAFGVYKGIWIPDEYKIPVAIKVIRENTSPKANEIIL 775
Db 703 AHLRIKETEFKKIKVLSGAFGVYKGLMIPDEKVKIKVIAKEIREATSPKANEIIL 762
QY 776 EAYVMAVGSPYVGRLLIGICTLSTVOLVLTQMPYGLLDHVRNKGRLGSOGLNNQMOI 835
Db 763 EAYVMAVDNPHVCRLLIGICTLSTVOLVLTQMPYGLLDHVRNKGRLGSOGLNNQMOI 822
QY 836 AKGSYLEDVRLVHRDLAARNVLKSPNHRKIDPGLARLLDIDEPYNAADGKVPITKM 895
Db 823 AKGNVYLEDRLVHRDLAARNVLKTPHVKITDFGLAKLGLAEEKEVYAEAGKVPITKM 882
QY 896 ALBSILRRFTHQSDVMSYGVYWEMLTFGAKPYDGI PAEILPDLLEKERRLPPICTI 955
Db 883 ALBSILHRIYTHQSDVMSYGVYWEMLTFGSKPYDGI PAEISILLEKERRLPPICTI 942
QY 956 DVYVIMVCKMNISECPRRRELVSSESNARBPQCFVYIQ-NEDGSPSPDSTFYNSL 1014
Db 943 DVYVIMVCKMNIIDSDSPKRRRLLESKARBPQCVLYIQGERHMLDPSPTDSFYRAL 1002
QY 1015 LEDDDMGDLVDAEBEYLVPOGFCPPAPGAGGVHRRSSSTRGGGDLTLGLBPSSE 1074
Db 1003 MEEDMEDVDADAEYLLPQGGF-----NSPST----- 1030

QY 1075 EAPRSLAPSBGAGSDVFDGDLGMAKGLQSLPTHPSPLOQYSEDPYPLPSET--DG 11
Db 1031 --SRTPLLSSISANSN-----SSTVACTINRNGSCVFKEDAFLOQYSDDPTVLTEDNIDPT 106
QY 1133 YVAPLTCSPOPEYVNPDPVPQPPSPREGPLPAAPAPAGATLERAKTUSPGKNGVXDVFA 111
Db 1085 FL-----FVPEYINQ--SVPRKPAQSGVQNPVYHNQPLHP-----APGRDLHYQN--P 114
QY 1193 FGAVAVENEYI--TQCGAAPQHPHPFPSPAFDNLVYWDQ-----DP-----PE 121
Db 1128 HSNMANSNPEYINTAQ-----FTCLSSGFSFSSALMIGKSHQMSLDNPDYQODFFPK 117
QY 1236 RGAPSTFKGPTAENPEYICLDVP 1260
Db 1179 EAKPVGIFKG-PTAENAEYLRVAP 1202

RESULT 3
Q9EP98
AC Q9EP98 PRELIMINARY; PRT; 1210 AA.
DT 01-MAR-2001 (TRENBLREL.16, Created)
DT 01-MAR-2001 (TRENBLREL.16, Last sequence update)
DT 01-JUN-2002 (TRENBLREL.21, Last annotation update)
DE Epidermal growth factor receptor isoform 1.
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVETAC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearraill R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maile N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Schenl C., Pearraill R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maile N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF275366; AAC28045.1; -.
DR EMBL; AF275364; AAC28045.1; JOINED.
DR EMBL; AF275365; AAC28045.1; JOINED.
DR EMBL; AF275367; AAC24386.1; -.
DR HSSP; P11362; IFGK.
DR MGJ; MGJ:95294; Egfr.
DR InterPro; IPR000345; Egfr.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_csr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU_5.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TYKc; 1.

```

DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN_1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP_1.
 DR PROSITE: PS00011; PROTEIN_KINASE_DOM_1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TIR_1.
 DR ATP-binding; Receptor; transferase.
 SQ SEQUENCE 1210 AA; 134840 MW; 62CD021C9DE32E18 CRC64;

Query Match 44.8%; Score 3073; DB 11; Length 1210;
 Best Local Similarity 49.2%; Pred No. 26-223;
 Matches 630; Conservative 168; Mismatches 362; Indels 120; Gaps 26;

QY 11 LLLALLPGAA--STQVCTGTDMKLELPASPETHLMDLRHLYOGCQVYQNLLEUTYPTN 68
 DB 14 LLTALCAAGALEEKVCOGTSNRLTQGTGFEDHLSLQRMNCEVYLGNETITVQRN 73
 QY 69 ASLSLELDIOEVQVYLIAHQVQVPLQRLIYVNGTOLFEDNALATLDNGDPLNTP 128
 DB 74 YDLSFLKTIQEVAGVYLAINTVERIPLENDQILKGNALYEVTALALISN----- 124
 QY 129 VTGASPDGLREQLRSLEILKGVYLIQNPOLCYQDTILWKDI---FHKNQALATLI 184
 DB 125 -YGTNRGTRELPMRLQEIILIGAVRFNNPILCNMDTIQWRDIQVNVFMSNMDL--- 180
 QY 185 DTRSRACHPCSPMCKSGRCWGESSEDOSLTRYACAGCA-RCKGRLPTCCGECQAG 243
 DB 181 -QSHPSCKPCDPSGSCNGSCWGGEGENCOKLTKIICAOQCGRGRGSPDCCNQCAG 239
 QY 244 CTGPKHSDCLACLFHNSGICELHCPALVTYNTDFEEMPRDEGRYTPGASCVTACPYNY 303
 DB 240 CTGPRRESCLVCGCKPQDEATCKDTPRLMLNPTTYQMDVNPBGKATCGATCVKCKRNY 299
 QY 304 LSTDVGSCTLVCPRLNQEYTAEDTQRCCKSKPCARVCYGLM--FNNFTVS-WLBPVK 361
 DB 300 VYTDHGSQVRACGPDYEVV-EDDG-RKCKKCDGPRKQCNIGIGEFKD-TLS----- 350
 QY 362 VSASHLEEFAGCKKIFGSLAFLPESFDGDPASNTAPQYIKANSFKITGTELTGLYISAM 421
 DB 351 INNTNINIKFKYCTAISGLDHLIPVAFKDSPTKTPRLDRELIELTKYKELTGLLIQAM 410
 QY 422 PSLPDLVSFQNLQVIRGRILHNGAVSLTQGLGISMGLRSLELGSGLALIHNTLHC 481
 DB 411 PDMWTDLHAFEMLEIIRGTRKHOGFSLAVVGLNITSLGLSLKEISGDVITISGNNTLC 470
 QY 482 FHTVPMQDLFBNPQALHTANREDECVGEGLACHQLCARGCKMGPRQCNCQFL 541
 DB 471 YANTIMWKKLFGTEWOKTKIMNRRAEKCKAVNHVCPNLCSSBGCWGEPRDVCSCNVIS 530
 QY 542 RQCEVCECVLQGLPREYVNAHGLPCHPRECPONGSVTCFGEADQCVACAHYKDPF 601
 DB 531 RRECEVEKCNILEGPREFEVENSECICQHPCLPQAMNITCTGSGPDMCICAHYIDGPH 590
 QY 602 CVARCSGVKPDLSIMPKFDEGACQPCPINTGHCVDLDKCGPAEGRASPLTSIV 661
 DB 591 CVKTCFAGIMGENNTL-VKTYADANNVCHLGHANTYACAGPGLQGEVWPSGKPIISA 649
 QY 662 SAVVGLLVVVLGVVFGI-LIKERQOKIKRYTMRLLQETELVEPLTPSGAMPQAOMRI 720
 DB 650 TGIYGLPLFIV-VALGIGLFMRHRHRIYRKRTLRLLQERLVEPLTPSGAPRQALIRI 708
 QY 721 LKTELAKYKVLGSAFGTVYKGIITPRGENVKIPVAIKVRENTSPKANKEIDEAAYM 780
 DB 709 LKTEFEKKIKVIGSAGFVYKGLMIPGEGYKIPVAIKELREKATSPKANKEIDEAAYM 768
 QY 781 AGVSPVYSRLIGICTSTVQLVTQMPYGLLDHVRNRRGLSODLLMCMQIAKMS 840
 DB 769 ASVDNPHVGRLLIGICTSTVQLITLMYGLLDYVRHKNITISQYLLMNCQIAKGMN 828
 QY 841 YLEDRVLRHDLAARNVLVKSBNHYKITDFGLARLLIDETEVHADGKYPIKMALESI 900
 DB 829 YLEDRVLRHDLAARNVLVKTPOHYKITDFGLARLLIGAEKEVYHAEGKYPIKMALESI 888
 QY 901 LRRFTHOSDVMSYVATWELMTFGAKRYDGIIPAREIPDLLEKERRLPPEPCTIDVYMI 960
 DB 901 LRRFTHOSDVMSYVATWELMTFGAKRYDGIIPAREIPDLLEKERRLPPEPCTIDVYMI 960

DB 889 LHRITHOSDVMSYVATWELMTFGSKRYDGIIPASDISLLEKGERLPPPEPCTIDVYMI 941
 QY 961 MKCKMIDSECRPRRELVSFSSMAARDPQGFVYIQ-NEDLGASPLDSTFYNSLEDD 10:
 DB 949 MKCKMIDASRPKRELILFESSKARDPQGFVYIQGDERHMLPSPDNSFYRALMDEED 10:
 QY 1020 MGDVDAEYVLVPOGFGFCPPDPAGAGMHHRRSSSTSGGDDLGLGEPSEEARPS 10:
 DB 1009 MEDVDADAEYLTPOGFGF-----NSPST-----SRT 10:
 QY 1080 PLAPSEGASDVFDGDLGMAKGLQS-LPTHDPSPRLQRYSEDPVPLPSET--DGYVAPL 11:
 DB 1035 PLSSLSATSN---NSTVACINRNGSCRYKEDAFLOQYSSDPFGAVTEDNIDDAFL-- 10:
 QY 1138 TCSPPQEVYNQDPVPPQPSRREGFLPAPRAGATLEBAKTLSPGKGVYKDVAFGAV 11:
 DB 1088 ---PVEEYVQ-SVPRKAGSVQNPVYHNPDLHP-----AGRDLDHYON--PHSNV 11:
 QY 1198 ENPEYL-TPGGGAAPQPPAPFSPAFDNLTYMDQ-----DP-----PERGAP 12:
 DB 1134 GNPEYLNTAQ-----PCLSSGFRSPALMTQKSHQMSLDNPQYQDFFPKETKPN 11:
 QY 1241 STPKGTPTAENPEYGLDVP 1260
 DB 1185 GIFKG-PTAENAEYLRAVP 1203

RESULT 4

QYH40 PRELIMINARY; PRT; 1165 AA.

AC QYH40; 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Receptor tyrosine kinase proto-oncogene.
 GN XMRK.
 OS Xiphophorus xiphidium.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorphi; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OX NCBI_TaxId=6086;
 RN [1] SEQUENCE FROM N.A.
 RP STRAIN=RIO PURIFICATION;
 RC MEDLINE=98241172; PubMed=9582016;
 RA Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,
 RT "Activation of the Xmrk proto-oncogene of Xiphophorus by
 RT overexpression and mutational alterations.",
 RL Oncogene 16:1681-1690(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIO PURIFICATION;
 RA Schartl M.,
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U53471; AAD10500.2; -.
 DR HSSP: P11362; IFGK.
 DR InterPro: IPR000345; CytC_heme_bind.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR InterPro: IPR001245; Tyr_kinase.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; Kinase; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR PRINTS: PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk_kinase; 1.
 DR SMART: SM00261; Fu; 3.
 DR SMART: SM00219; TYRK; 1.
 DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN_1.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.

DR PROSITE; PS00107; PROTEIN KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN KINASE_DOM; 2.
 DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
 KW ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 1165 AA; 129614 MW; 757EB3BD8771A74E CRC64;

Query Match 39.2%; Score 2688; DB 13; Length 1165;
 Best Local Similarity 45.0%; Pred. No. 2,66-194;
 Matches 575; Conservative 172; Mismatches 384; Indels 148; Gaps 31;

1 MELALCRMGLLMLPRG-AAST-----QVCTGDMRLRLPASETHLMDLRLYGGCCV 55
 4 LELLE-----LLLLLSIGRCCSTDPDRKVCQGSNQM---LDNYYLKKMKMYSGCAV 56
 56 VQNLLELYPTNLSLFLDIOEVGVYLIANHOVQVPLQRLRIVGTQLFEDYVALA 115
 57 VLENLEITYQENDLSFLQSIQEVGVYLIANNEVETIPLVNLRLRQNLKVEGFTLL 116
 116 VLDNDPLANTTPTVTGASPGRLRLQRLSLTELKGVLIQRNPQLCYDTILMKDIFRK 175
 117 VMSNYQR-NPSSP--DYVQVGLKQLSLNLTLEILSGGVKSHNPLLCNVEITIMMDIVDK 173
 176 NNOLALTLIDTNRBRACHPCSPMCKSGRCWSESESDCQSLTRVYCAAGC-ARCKGFLPTD 234
 174 TSNPTNMLIPHAFFRQCQKDCDPCGVNCSWAPRGHCKFTKLCAEQCNRRRGKPID 233
 235 CCEHQAAGCTGPHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNDEGRYTFGAS 294
 234 CCNEHAGAGCTGPRATOCACRDPNDGTCDCDTPPKIYDVSHQVYDNPNIKYTFGAA 293
 295 CVTACPVNYLTDVGSCTLVPLNCEVTAEDGQREKSKCAVAVCYLGMFN-NFTV 353
 294 CVKECPNIVYTE-GACVRCSSAGMLEVD-ENGRSKCPDGVCPKVCDDIGISLSNTI 351
 354 SFWDRVPRVSASHLEPAGCKKIFGSLAFPESEPDGPASNTAPQYIKANSKFTGTEYT 413
 352 A-----VNSTNGSFENCTKINGDIILNRSPEGPHYKIGPMDPEHMLNLTYTEIT 404
 414 GYLIISAMPDLPPLSVFQNLQVIRGILNNGAVS-LTLOGLSIWLGLRLSELGSLA 472
 405 GYLVIMMPEMNTSLSVFQNLLEIRGRTFSRGSPFVVVSHLQMLGLSSLLEVSAGNV 464
 473 LIHNHTLCFHTVPMQDLFRNPQALLHTANRPEDECEVGEGLACHQLCARHCWMPGPT 532
 465 ILKNTPOLRVASTINMRRLFRSEDSQSIETDART-----ENQTCNNEGSEDDCWPGPT 517
 533 CCVNCQFLRGQCEVEECRVLQGLPREYVNAHCLPCHPECQPOQNSVTCFGEADQCYA 592
 518 MCVSCLHYDRGRCVASCNLIQGFREAOVDGRCVQCHQCELVQTDLSLCYGGPANCSK 577
 593 CAHYKDPFCVARGPSGVKPDLSYMPIMKFPDEEGACOPRINCTHSCVDLDKCGPABO 652
 578 CAHFQDGPQCTPRPHGLMDGDTL-TWKADAKKQGCOPHONCTGCGGPGISGGRG- 635
 653 RASPLTISVSAVVGILLVVLGVVFGILLIRROQKIRKYMRLLOETELVEPLTPSGAM 712
 636 IVSHSSLAVGLVSGLLITVIALIIVLLRRRIK-RKRTIRLLQEKELVEPLTPSGA 694
 713 PNOQMRILKETELRKXKULGSGAFGVYVYGMIPGSENIKIVALKVIRENTSPKANE 772
 695 PNOAFRLILKETELRKORVJSGGFGVYVGLMNPDEENIRIVALKVIREAISPWONE 754
 773 ILDEAYVAGVGSFVYSRLGICLTSTVQLVTQJLMPYGLLLDVARENRRLSGODILANC 832
 755 VLDEAYVAVSDHHCRLGICLTSAVQVLVTQJLMPYGLLDVVRQHORIGQMLINANC 814
 833 MQIAKMSYLEDVLLVNRDLAARVLYKSNPKYITDFGLARLLDIDETFEYHADGKVP 892
 815 VOIAKGNVLEERHLVARDLAARVLLKKNHAKITDFGSKLLTADKEKXQAHGKXFI 874
 893 KMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIDLEKGRLLPOPI 952
 875 KMALESILQMTYHQSDVMSYGVTVWELMTFGSKPYDGIIPAKELASVLENGERLPOPI 934

QY 953 CTIDVYMMVYKMMIDSECRPRELVSFSPRMAADPQRFVIONEDLGPASPLDSTFYR 101
 DB 935 CTIEVYMLILKMMIDSSRRFRFELVGEFSQMARDSRLVLIQG---NLPSYSDRLFS 991
 QY 1013 SLEDDMGDLVDAEYLVPOQGFPCDPAPAGAGVYHRRSSSTRSGGDLTLLGELS 107
 DB 992 RLSSSD--DVYDADEYLL-----RYKRN-RQGS----- 101
 QY 1073 EEEARPSILASEAGSDVDFDGLGMAAGLQSLPHDSPLQORSSEDTV-PLRSSETD 113
 DB 1019 -----ECIIPPNH-----PVRENSIALRYISDPTQNALEXDLD 105
 QY 1132 GYVAPLTCSPQPEYVNPDPVRRP-----PSPRE-----GPLP-AAPAGATLSEAKT 117
 DB 1053 GH-----EYVNPQGSSETSRSLSDINPVYEDLTQMGCVSLSSQCAETNFSRPEY 110
 QY 1179 LSPKNGVYKDVFAFGAVENPEYLTFGGAPAPQHPHPPAFSFAFDNLYWDDPPERGA 123
 DB 1103 LNTQNSL---PLVSSGSMDDPDY---QAG-----YQAAF-----LPQTGA 113
 QY 1239 PSTFTKGTPTAENPEYGL 1257
 DB 1138 LTGNGMFLPAENLEYGL 1156

RESULT 5

Q9MEF6 PRELIMINARY; PRT; 1137 AA.

AC Q9MEF6
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Receptor tyrosine kinase (Fragment).
 GN ERB4.
 OS Gallus gallus (chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=HINDRAIN;
 X MEDLINE=99263203; PubMed=10328884;
 RA Dixon M., Lumsden A.;
 RT "Distribution of neuregulin-1 (nrg1) and erb4 transcripts in
 RT embryonic chick hindbrain.";
 RL Mol. Cell. Neurosci. 13:237-258(1999).
 DR EMBL: AF121963; AAD31764.1; .
 DR HSSP; P11362.1FGK.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_Pkinase.
 DR InterPro; IPR002174; Furin-1like.
 DR InterPro; IPR001368; TNFR_c6.
 DR InterPro; IPR001245; Tyr_Pkinase.
 DR InterPro; IPR004019; YLP_motif.
 DR Pfam; PF00757; Furin-1like; 1.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF01030; Recep_L_domain; 1.
 DR Pfam; PF02757; YLP; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_Pkinase; 1.
 DR SMART; SM00261; FU; 3.
 DR SMART; SM00219; TYKc; 1.
 DR PROSITE; PS00107; PROTEIN KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
 DR PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.
 KW Kinase; Tyrosine-protein kinase.
 FT NON TER 1
 SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;

Query Match

38.8%; Score 2660.5; DB 13; Length 1137;

Best Local Similarity 46.4%; Pred. No. 3,1e-192;
Matches 531; Conservative 171; Mismatches 360; Indels 83; Gaps 25;

```

QY 161 LCYQDTILMKDI FHKNNQLALTITDNRSPACHPCSPCKSCRCWGESSEDQGLTETVC 220
DB 3 LCFADTIDHMODIVRNPMASNFTLVPTNNGSGGGRCHKSCG-RCWGPTEHNCQTLTETVC 61
QY 221 AGGC-ARCGPLPTDCCHEGCAAGCTGPRHSDCLA CLHFNHSGI CELHCPALVTYVNDTF 279
DB 62 AEGCGRCTGPIVSDCHREKAGCGSPDITDCAKMFNDGACVTCQPGTFYVNTTF 121
QY 280 ESMNPDEGRYTFGASCVTACPYVLTSTVGSCVTLVPLHNOEVTAEADGTCRCKSKPCA 339
DB 122 QLEHNHNAKYTYGAFCKKCPENFV-VDSSCVCRPSKMEV-ENKGIKKCKCTDICTP 179
QY 340 RVCYGLGEMNFTVFEMLRVPRKVSASHLEEFAGCKKIGSLAFIPESIDGPAANTPQY 399
DB 180 KACDGI-----TSLVSAQTVSDSNIDKFNCKIKNGNLIPLVTGIGHGDPYHTIAIN 233
QY 400 IKANSKFIGITELTGLYLSAMPDSLPLSVFQNLQVIRGRILHNGAVSLTQGLGISML 459
DB 234 PEKLNIFQTVREITGLVNIQSWPEKMTDFRVFNSNLVTIGRAIVSGSLILKQGITSL 293
QY 460 GLRSLRELSSGLALHNHTHLCFTVTVMDQLFRNPHALHTANRPEDCVGEGLAHQ 519
DB 294 OFOSLKQISAGNIVTIDNSNLCTVHTVMTSLFSTPSQKTIHKKKAENCTADGWCNE 353
QY 520 LCARGHGMGPPTQCNCSQFLRQGECEVECRVLQGLPREVYNAHCLPCHPECOP-QNG 578
DB 354 LSSDGCWGPDPDQLSCKRFIRGTCIESCNLYGFEREIRANSVCMECDPOCKEKEDN 413
QY 579 SVTCFGEADQCVACAHYKDPFCVACRPSGVKPDLSYMPIKFPDEGACQPCPINTH 638
DB 414 MTCYGPGRDHCTCKFHPKDGNCVEKCPDGLQGANSE--FFKXADEBDECHPCPNCTQ 471
QY 639 SCVDLDDKGC-----PAQRASPLTISAVV-GILLVYVIGVYGLIKRQ 666
DB 472 GGRGASHDCIYPTKROSTLPQNR-TPL--IAGVIGGFLTVINGLTFAYVVRKST 528
QY 687 KIRKYTMRLLOETELVBLTPSGAMPNQAQMLIKETELRKVVLGSGAGTVYKGIW 746
DB 529 K-KKRLRRFL-ETELVBLTPSGTAPQAQRLIKETELRKVVLGSGAGTVYKGIW 566
QY 747 PDGENVKI PVAIKVRENTSPKANELLDEAVYMAVGSPIVSLILGICLTSTQVLTOL 806
DB 587 PGETVKKLPVAIKLINETTPRANVEFDEALIMASMDHPVLVLGVCLSPITQVLTOL 646
QY 807 MPYGCULDHVENRGLSGODLLNMCQIAKMSYLEVDVRLVHDLAARVLYKSPHYK 866
DB 647 MHPGCLDLYHEHKNIGSQLLNMCVQIAKMMVLEBRRLVHRLDLAARVLYKSPHYK 706
QY 867 IINDGLARLLDIDELHYADGKVP IKMALESLLRRAFTHQSVKSYGVTVWELMTGA 926
DB 707 IYDFGLARLLEDEHYADGKVP IKMALESLHRKFTHQSVKSYGVTVWELMTFGG 766
QY 927 KEYDGI PAREIPLDLKGERLPOPICTIDYVIMVCMWIDSECRPFREIYSEFSRMA 966
DB 767 KEYDGI PAREIPLDLKGERLPOPICTIDYVIMVCMWIDSECRPFREIYSEFSRMA 826
QY 987 RDPQRFVITQNEQ-LGPA SPLDSTYRSLLEDDMDQIDVAEEVLYVQGFPCPDPAFGA 1045
DB 827 RDPQRFVITQNEQ-LGPA SPLDSTYRSLLEDDMDQIDVAEEVLYVQGFPCPDPAFGA 885
QY 1046 GGMVHRHRSSTSGGDLTLGLEPSESEAPRS--PLAP-SEGAGSDVDGDLGMAAK 1102
DB 886 RTRIDSNRNOFYVRGGYAAEGV-PMYRAPGCLIEBAPVQAATFEIEDCCNGTLR 944
QY 1103 GLQSLPTHDPBPLQRYSEDPTVPLPS-----ETDGVAPLTCSPPREVNGPDVAPQ 1155
DB 945 KQAVTAKEDSSTORYSADPTVFIPEKVI RGELEDEDEGYMTPMRKPKXTDYLNVEENPFV 1004
QY 1156 PSPREGPLPAA-RPAGATLEBAKTLSPCKNGVVDV-----AGCAVNEPEVLTITQ 1206

```

```

DB 1005 SRKKNGD:QAVDNPENYH-----APNGCAPKADDEVYNEPLVINTFANTLENAYL--- 107
QY 1207 GGADPQHPPEPPAFSFAEDNLYWDQDPPERGA--PSTFKGTPT-----AE 121
DB 1055 -----KNLPEKAKKAFNDPDMYHNSLPPRSTLQHPDVLQYSTKYFGKNGRIRPIAE 114
QY 1251 NPEYL 1255
DB 1110 NPEYL 1114

```

RESULT 6

```

ID P79754 PRELIMINARY; PRT; 1328 AA.
AC P79754;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE ErbB3.
US Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
NCBI_Taxid=31033;
RN [1]
RP SEQUENCE FROM N. A.
RX MEDLINE=99177477; PubMed=10077531;
RT Cellner K., Bremner S.,
RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
RT rubripes."
RL Genome Res. 9:251-258(1999).
DR EMBL; AF056116; AAC34391.1; -
DR HSP; P11362; IFGK
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATF; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ
SEQUENCE 1328 AA; 148613 MW; A333039258B647B9 CRC64;

```

Query Match 33.3%; Score 2280; DB 13; Length 1328;

Best Local Similarity 40.1%; Pred. No. 2,4e-163;

```

Matches 518; Conservative 153; Mismatches 425; Indels 196; Gaps 31
QY 9 WGLLALLP--GAATG---VCTGDMKLRLPASPEHLDMRLHLYOCQVQVGNLEL 62
DB 4 WRLIMCVASLRPAASQGTQAVPGTQNGLSSTGSENGNYNLKORYKGCETIMGLET 63
QY 63 TYLPTNASLSLFDIGEOGVYLLAHNQVQVPLQRLVIRGTQLFEDNYALAVLDNGDP 122
DB 64 TQIESNMDFSLKTRREVTGVLIAHMHFQELPGLRLVIRGNSLVERRALSVFLN--- 120
QY 123 LNNTPTVGAAPGLRELO:RSLTLLKGGVLLQRPQLCYOTITLMKQIFHKNNQLALT 162
DB 121 ----YPKDG--PSGLNQLGNLLEIILDGAVQIINNYLRYGFWYWRDII-RNNDAPIE 173
QY 183 LIDNRSRACHPCSPMKGRCWGESSEDQSLTRVCAAGC-ARCGPLPTDCCHEGCA 241
DB 174 IQFNIGRGVCH---KSC-GNYCWGPGKQDQQLITKYCAQCNDRCFGSPRCCHIEA 229
QY 242 AGCGPRKHSDDLALHNHNSGICELHCPALVTYVNTDIFESMPYDEGRYTFGASCVTACY 301
DB 230 AGCGPLDITDCAFLRPNDSGACVPOCPQTLINVKQTFQMETPNPAKYGSGISVSCPT 289

```

```

QY 302 NYLSTDVSGCTLVCPFLHNOEYTAEDGTOR-CEKSKSCPCARVYCYGLGFNNFTVSWFLWLP 360
DB 290 HFV-VDSSCVSVCPDMEV--ERSGRQCELCESGCPKCESTGABQGT----- 338
QY 361 KVSASHLEEPACCKITPSSLAFLPESFPDPASTAPQYIKANSKFIGITELGYLISA 420
DB 339 -VDSSNIDSFNCKTIQSLHFLTGLTGDDFKVPPIDAKKEVFRTVEIDIDINTIOS 397
QY 421 WPDSPDLSVFQNLQVIRGRILHNGAVSLTQGLGIMLGRSLREFSGSGLIHNHTH 480
DB 398 WPKELNDLSVFSLSLTIGRSLFRKFSLMWRIPTLISGLRSIRESDSVISQNAHL 457
QY 481 CFVTFVWDQLEFRNH-QALLHTANRPDESCVGEGLACHQICARHGCHWGPPTQVNCQ 539
DB 458 CYHHTVNWTLFRGSRVANSLSNRPMAECVADGRVCDPLCSGSGCWGPGDQCLSCRN 517
QY 540 FLRQGECEVECKVLOGLPREYVNAH-GLPCHPCQONSVCFCGEADOCVACAHYD 598
DB 518 YSRHGTCAAGCHFNHSGIPREFAGLNGVCVACHPECKPOTGKASTGTGADBCMACTKFRD 577
QY 599 PPECVARCPGSKPDLSTYMPIWKEPDEEGACQPCPINCSTHSCVDLDDKGPASQAPSLT 658
DB 578 GPYCMSSCPAGVN-DGEKGLIFKEPNEGHEPCHQNCTQCCSGPGLNDC--LEAARLT 633
QY 659 STVSAVVGILLVVLGVF-----GILLRQOKIRKXTMRLLDETLEVPILPSGA 711
DB 634 ISSGQITGIALGVPAGLIFCLVLFELGVLHRLGALIRKRAMRYLESGBEPFLP-GE 692
QY 712 MPNQAQRILKETELRKVVLGSAFGTVYKGIWIPGENVTKIPVAIKVIRENTSPANK 771
DB 693 KGTXYHARILKPSDLRIKPLGSGVFGTVSGKPMIPGEYTKIPVALKITQDSSGQGTFT 752
QY 772 EILDEAYMAGVGSYSYRLIGITCTSTVQLVTLQMLYGCCLDHVRENRGLSQDLWM 831
DB 753 EITDILSMGSLDHPYIVRLIGIPGCLQVLTQSSHGSLLEIRHOKTSLPQBLWM 812
QY 832 CMOIAKMSYLEDEVLVHRDLAARNVLVKSNNHYKINDPGLARLDIDEVHADGKVP 891
DB 813 CVOIAKMGVYLEHRVYHKNLAARNILKNDYQVQISDYVADLLPDDKKVYVSEKTP 872
QY 892 IKMALESILRRRTTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLKGER-POPP 951
DB 873 IKMALESILRRRTTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLKGER-POPP 932
QY 952 ICTIDVYIMWKCMWIDSECPREFELVSESRMAPDQCFPVITQNEIDLGPASPLDSTY 1011
DB 933 ICTIDVYIMWKCMWIDSECPREFELVSESRMAPDQCFPVITQNEIDLGPASPLDSTY 980
QY 1012 RSLLEDDMDGLVDAEYLYVPOGFCPPDAPAGAGVWHHRSSSTRSGGDLTLGLEP 1071
DB 981 ----EDSGMGFL-----RGSER--GILLEADLEB 1004
QY 1072 SEEPAPRSPLAPSEGAGSDVFDGLMG--AAKGLOSPTHPSPLQ-----RY 1118
DB 1005 DEER-----GLGDRFATPSLOPSPSWSTSPQINSYVWMTQLRY 1043
QY 1119 SEDPTVPLPSETDGVAPLTCSPOP-EYVNO-----PVYRQPSPREGL 1163
DB 1044 D-----FAVSQGHIGYLPMSBPVVTIRQLMYORSLSSVATLPRRAFRRSSAEAL 1097
QY 1164 --PAARPAGATLEAKTSLPEKGNVAVDPAFGAVENPEYLLTPQGAAPQHPAPFSP 1221
DB 1098 CEDGAQCGIFRVE-----FGSERGN-----PQGG----- 1122
QY 1222 AFDNLVYWDODPPERGAPSPSTFKGTFAENPE 1253
DB 1123 -----QQRKLSSTASPSFKTWADDEDE 1146

```

RESULT 7
Q9BIH9 PRELIMINARY; PRT; 1433 AA.
AC Q9BIH9;

```

DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles
OX NCBI_Taxid:7165;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=SU4;
RA "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301655; CAC35008.1; -.
DR HSP; P11362; IFGR.
DR InterPro; IPR000345; CyC_heme_bind.
DR InterPro; IPR004954; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Dkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PS00109; TYRKINASE.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKc; 1.
DR PROSITE; PS001219; TyKc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
FT Receptor.
FT NON_TER.
SQ
SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;
Query Match 29.4%; Score 2011.5; DB:5; Length 1433;
Beet Local Similarity 32.3%; Pred.No.576-143;
Matches 469; Conservative 197; Mismatches 396; Indels 391; Gaps 37
QY 26 CTGDMKRLRLPASPETHLDMRLHYQSCQVWQNLBELTYPTVASLFIODIOEGVYVL 85
DB 1 CIGTNGRMSVARNREYHYKNLRDXYTNCITVDGSLLETWQNTIDNLFQIHIREVGYVL 60
QY 86 IAHQVQVPLQRIKRYRGTLF-----EDNYALAVLDNGDPLNNTTPVTGASPGGLREL 140
DB 61 ISLYDLPGVILPRQIIRGRTTFGLNKEBAYGLFV-----SFSHMNTL 104
QY 141 QLRSLTELKQGVILQRPQICVQDITLWMDI-FHKNQNLATLIDNRSRACHPSPMC 199
DB 105 ELPRALRDLIGSGVGFNNYVNLCHKNSINWEILLAPQTSQYTFNFSBPVRVPCHPSC 164
QY 200 KGRSCWGESSEDCQSLRTVYCAQCA--RCKGPLPTDCCHEQCAQCTGPKSKDCLACLH 257
DB 165 EVG-CWGEAGHNCQRFKSLNCSQCSGRCFGKPRKPCCHLFCAGCTGTGTDCLACKN 223
QY 258 FNHSGICELHCPRLVTVNTDTFESMPREBERYTFGASCTACGYNTLSIDVSGCTIVCL 317
DB 224 FYDQGVCKQCPRPQIYNPTNYFWEPRPDQKAYGATCVAKCP-EHLKONGACVARKCP 282
QY 318 HNOETVADGTORCEKSKPCARVYCYGLGFNNFTVSWFLWLPKVSASHLEEPACCKITF 377
DB 283 GRMPQNSE-----CVPCKGVCPKTCGEGE-----VASDNYGNKQCTIIE 323
QY 378 GSLLAFLESPFDGPASNT---APQYIKAN-----SKFIGITELGYLISANPDSLPDIS 429
DB 324 GSLEILDQSPDGRQOVYTNFSFSPRYIKIDIPRLLEVFSVKEITGFINIQAHHPNFTLIN 383

```


RN (1)
 Q86712 SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC027080; AAH27080.1;
 SK Hypothetical protein.
 SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E10180 CRC64;

Query Match 25.4%; Score 1739; DB 11; Length 367;
 Best Local Similarity 88.0%; Pred. No. 3,1e-123;
 Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 895 MALESLIRRTTHOSDWSYGVYTWELMTFGAKYDGIIPAREIPDLLEKGERLPDPIC 954
 Db 1 MALESILRRRTTHOSDWSYGVYTWELMTFGAKYDGIIPAREIPDLLEKGERLPDPIC 60
 QY 955 IDVYIMYKCMWIDSECRPRFRELVSERPMARPCGFVVIQNEEDLGPASPLSTYRSI 1014
 Db 61 IDVYIMYKCMWIDSECRPRFRELVSERPMARPCGFVVIQNEEDLGPASPLSTYRSI 120
 QY 1015 LEDDDMGDLVDAEYLVPQCGFCPPDPAFGAGMVRHRRSSSTRSGGDLTLGLEPSEE 1074
 Db 121 LEDDDMGDLVDAEYLVPQCGFCPPDPAFGAGMVRHRRSSSTRSGGDLTLGLEPSEE 180
 QY 1075 EAPRSPLAPSEAGSDVFDGDLGMAKGLQSLPTHDPSPLORYSEDPYLPLESETDGY 1134
 Db 181 EPPRSPLAPSEAGSDVFDGDLAVGVTKGLQSLPHDLSPLQRYSEDPYLPLESETDGY 240
 QY 1135 APLTCSPOPEYVNPQDVPRPSPREBGLPAPAREGATLERAKTLSPGKNGVYKDFAFG 1194
 Db 241 APLACSPQPEYVNPQDVPRPSPREBGLPAPAREGATLERAKTLSPGKNGVYKDFAFG 300
 QY 1195 GAVENPEYLTQGGAAQPHPPAFSPAFDNLVYWDQDPPERGAPOSTFKGTATANPEY 1254
 Db 301 GAVENPEYLTARAGTASQPHPSPAFSPAFDNLVYWDQDPPERGAPOSTFKGTATANPEY 360
 QY 1255 LGLDVPV 1261
 Db 361 LGLDVPV 367

RESULT 10
 Q86712 PRELIMINARY; PRT; 729 AA.

AC Q86712;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE Polypeptide.
 GN POLYPEPTIDE.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
 CX NCB1_TaxID=11950;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; Pubmed=8152791;
 RA Vennistrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
 RA Johnson A., Beug H.;
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities.";
 RL Oncogene 9:1307-1320(1994).
 DR EMBL: S69372; AAC60725.1; -;
 DR HSP; P03322; IAGS.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR004028; Retro_M.
 DR InterPro: IPR001245; Tyr_kinase.
 DR Pfam: PF00069; Pkinase; 1.
 DR Pfam: PF02813; Retro_M; 1.
 DR ProDom: PD000001; Euk_kinase; 1.
 DR SMART: SM00219; TYRK; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PSS0011; PROTEIN_KINASE_DOM; 1.

DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 729 AA; 80649 MW; 84D26914FE1D53 CRC64;

Query Match 25.1%; Score 1720; DB 15; Length 729;
 Best Local Similarity 54.8%; Pred. No. 2.4e-121;
 Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 1;

QY 575 PONGSVTCFGEADQCVACHYKDPRECVARCPBGVYKPDLSYPMIMFPEDEGACPCPI 634
 Db 141 PEETATPTTGP--DHCKCAHFIDGPHCVACPAVGENDTL-VMKYADANAACVQLCHP 197
 QY 635 NCTHSCVDLDKCGCPACORASPLTISYSAVY-GILVVVGVGVGILIKRQCKIRKYM 693
 Db 198 NCTGCKCPGLEGCP---NSKTPSIAGAVGVGILCVVGVGLGILYLR--HIVKRTI 253
 QY 694 RRLIQTETLVPLPLPSGAMPNOAMRLKTELRKYVYLSGAFYVYKIMIPDGENVK 753
 Db 254 RRLIQERLVEPLPLPSGAPNOALRLKTELRKYVYLSGAFYVYKIMIPDGENVK 313
 QY 754 IPVAIKVLRNTSPKANKELIDEAYVWAGVSPYVSLGICLTSTYQVLTOLMPYGCIL 813
 Db 314 IPVAIKELRENTSPKANKELIDEAYVWASVDNRVCHLGLCLSTYQVLTOLMPYGCIL 373
 QY 814 DHEENRGRLGSDQLNMCQIAKMSYLEDVRLVHRDLAARNVLYKSPNRYKXTDFGLA 873
 Db 374 DYREHKDNISQYLLNMCVQIAKMSYLEDVRLVHRDLAARNVLYKSPNRYKXTDFGLA 433
 QY 874 RLIDIDETEVYADOGKXPIKMALESLIRRTTHOSDWSYGVYTWELMTFGAKYDGI 933
 Db 434 KLLGADAEKETHABGKXPIKMALESLIRRTTHOSDWSYGVYTWELMTFGAKYDGI 493
 QY 934 AREIPDLLEKGERLPDPICITIDVYIMYKCMWIDSECRPRFRELVSERPMARPCGF 993
 Db 494 ASEISVYLEKGERLPDPICITIDVYIMYKCMWIDSECRPRFRELVSERPMARPCGF 553
 QY 994 VIO-NEDLGASPLDSTFTYRSLLEDMDGDLVDAEYLVPQCGFCPPDPAFGAGMVRH 105
 Db 554 VIQDERMHLPSPDTSKFRYRLTMEEDMEDIVDADEYLVPHQGF----- 598
 QY 1053 HRSSSTRSGGDLTLGLEPSEEPAPRSPPL-----APSEAGSDVFDGDLGMAKGLQSL 110
 Db 599 -NSPT-----SRPFLSSLSATSNNSATNCID-----RNGQGH 631
 QY 1108 PTHDPSPLQRYSEDPYLPSET--DGYAAPLTCSPQPEYVNPQDVPRPSPREBGLPA 116
 Db 632 PVREDSFVQRYSSDPNGLFLESIDGFL-----PAPEYVNP--LMPKKS----- 675
 QY 1166 ARPAGATLERAKTLSPGKNGVYKDFV-----AFGAVENPEYL 1203
 Db 676 -----TANVQNIYNNISLTAISKLPMDSRYNQSHSTAVDNPXYL 715

RESULT 11

Q86714 PRELIMINARY; PRT; 567 AA.

AC Q86714;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE V-erbB protein (Fragment).
 GN V-ERBB.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
 CX NCB1_TaxID=11950;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; Pubmed=8152791;
 RA Vennistrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
 RA Johnson A., Beug H.;
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities.";

```

RL  Oncogene 9:1307-1320(1994).
DR  EMBL; S69372; AAC60727.1; -.
DR  HSSP; P11362; 1FGK.
DR  InterPro; IPR000719; Euk_pkinase.
DR  InterPro; IPR001245; Tyr_pkinase.
DR  Pfam; PF00069; pkinase; 1.
DR  PRINTS; PR00109; TYRKINASE.
DR  ProDom; PD000001; Euk_pkinase; 1.
DR  SMART; SM00219; Tyrc; 1.
DR  PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR  PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR  PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KM  Tyrosine-protein kinase.
FT  NON_TER
SQ  SEQUENCE 567 AA; 63390 MW; C6D9CBA7AD7F25E1 CRC64;

Query Match      25.1%; Score 1718; DB 15; Length 567;
Best Local Similarity 55.4%; Pred. No. 2,3e-121;
Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY  584 GPEADQCAAHYKDPFCVACSPGVKPDLSYPIWKPEDEGACQPCPINCTHSCVDL
DB  1 GP--DHCMCAHFIDGPHCVACAFAGVLGENDTL-VMKYADANAVCQLCHNCTRGCKGP
DB  644 DDGCPAEGRAPELSTISAVV-GILVVVLGVVFGILIKRQCKIRKYMRLLOTEL
DB  58 GLEGP--NSKRTSIAAGVVGGLCLVVGGLGILYLR--HIVKRTLRLLOREL
QY  703 VEPLTPSGAMPNQAQMLIKETELERKVLGSGAFGVYKGIWIPDENYKIPAIKVLK
DB  114 VEPLTPSGEAPQAHLRIKETEFKVKVLSGAFGVYKGLWIPDEKXKIPAIKELR
QY  763 EHTSPKANEILDEYVWAGSPVYSRLGICLTSTYQVLTQMPRGCLLDHYRENRGR
DB  174 EHTSPKANEILDEYVWASVDNPRVCLGICLTSTYQVLTQMPRGCLLDYIREKDN
QY  823 LGSQDLNMCQIAKMGSYLEDVRLVHRDLAARVLYKSPKHVXITDPFGARLLIDETE
DB  234 IGSQYLLNMCQIAKMGVYLEDERRLVHRDLAARVLYKTPHVKITDPFGAKLLGADKE
QY  883 YHADGKVPITKMALESILRRRFTHQSDVSYGVTVWELMTFGAKPYDGIPAREIPDLLE
DB  294 YHAEKGKVPITKMALESILRRYTHQSDVSYGVTVWELMTFGAKPYDGIPAREISSYLE
QY  943 KGERLPQPICTIDVYIMWKCMIMIDSECRPRELSEFSRMAQDQRRVVO-NEDLG 1001
DB  354 KGERLPQPICTIDVYIMWKCMIMIDADSRKRELIAEFSKQARDPRLYLQGDGRWH 413
QY  1002 PASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFPCPDPAFGAGVHHRRSSSTRSG
DB  414 LPSPTDSKFRYTLMEEDMEDIVDAEYLVHQGF-----NSPST--- 454
QY  1062 GGDULTLGLPESEEEFRSPL-----APSEGASVVFQGDGLGMAKGLQLPTPHDSPLQ 1116
DB  455 -----SKTPLLSLSATSNNSATNCID-----RNGQGHVRESQFVQ 491
QY  1117 RYSEDPFVLPSET--DGYVAPLTCSPQPEYVNOPDVRPOPSPREGLPLAARPAATLE
DB  492 RYSDPFGNPLEESIDQGL-----PAPEYVVO--LMPKKPS----- 1174
QY  1175 RAKTISPGKGVVDVF-----AAGGAVENPEYL 1203
DB  527 ----TAMVQNOIYNNISITAIKSLPKMSRYONSHSTAVDPEYL 566

```

```

GN  P31659.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX  NCB1_TaxId=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
RA  Man D.F., Gu J.R.;
RT  "Novel human cDNA clones with function of inhibiting cancer cell
RT  growth."
RL  Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AF318349; AAL5856.1; -.
DR  InterPro; IPR002048; EF-hand.
DR  InterPro; IPR000719; Euk_pkinase.
DR  InterPro; IPR001245; Tyr_pkinase.
DR  InterPro; IPR004019; YLP_motif.
DR  Pfam; PF00069; pkinase; 1.
DR  Pfam; PF02757; YLP; 2.
DR  ProDom; PD000001; Euk_pkinase; 1.
DR  SMART; SM00219; Tyrc; 1.
DR  PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR  PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KM  Hypothetical protein.
SQ  SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match      24.8%; Score 1697.5; DB 4; Length 412;
Best Local Similarity 80.5%; Pred. No. 5.1e-120;
Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4

QY  895 MALSELRRRTHQSDVWSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPICT
DB  1 MALSELRRRTHQSDVWSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPICT 60
QY  955 IDVYIMWKCMIMIDSECRPRELSEFSRMAQDQRRVVO-NEDLGAPSLDSTFYRSL
DB  61 IDVYIMWKCMIMIDSECRPRELSEFSRMAQDQRRVVO-NEDLGAPSLDSTFYRSL 120
QY  1015 LEDDDMDGLVDAEYLVPOQGFPCPDPAFGAGVHHRRSSSTRSGGDLTGLPSESE
DB  121 LEDDDMDGLVDAEYLVPOQGFPCPDPAFGAGVHHRRSSSTRSGGDLTGLPSESE 180
QY  1075 EAPSPPLAPSGASDVFDGLGMAKGLQSLPTHDPSLQYSEDPVPLPSETDGYV
DB  181 EAPSPPLAPSGASDVFDGLGMAKGLQSLPTHDPSLQYSEDPVPLPSETDGYV 240
QY  1135 APLTCSPQPEYVNOPDVRPOPSPREGLPLAARPAATLERAKTISPGKGVVDVFAPG
DB  241 APLTCSPQPEYVNOPDVRPOPSPREGLPLAARPAATLERAKTISPGKGVVDVFAPG 300
QY  1195 GAVENPEYLTPOGGAAPQ-----HPPPA--FSPAFLNL 122
DB  301 GAVENPEYLTPOGGAAPLTLPSAQPTTSITGRTHQSGGLHAPSGHLRQRTOST 360
QY  1227 YYMD-QDPPER-----GAPSTFGKTPTAEN 1251
DB  361 WWTCCQCEPQGRKSPDVSSGREGILTSAGIKRWECPTTSGTCHARN 410

```

```

RESULT 12
08WYVO PRELIMINARY; PRT; 412 AA.
AC 08WYVO.
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 44.7 kDa protein.

```

```

RESULT 13
064895 PRELIMINARY; PRT; 962 AA.
AC 064895.
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Gag-V-erb-A-V-erb-B protein.
GN GAG-V-ERB-A-V-ERB-B.
OS Avian erythroblastosis virus.
OC Viruses; Retrod viruses; Retroviridae; Avian type C retroviruses.
OX NCB1_TaxId=11861;
RN [1]
RP SEQUENCE FROM N.A.

```

RX MEDLINE=90206603; PubMed=1969616;
 RA Biskin A., Jackson J., Bishop J.M., McCarty D.J., Schatzman R.C.;
 RT "Six amino acids from the retroviral gene gag greatly enhance the
 RT transforming potential of the oncogene v-erb-B.";
 RT Oncogene 5:15-24(1990)
 CC - SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC - SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 DR EMBL; X52209; CA36459.1; -.
 DR EMBL; X52211; CA36459.1; JOINED.
 DR HSSP; P10828; 2NL.
 DR InterPro; IPR000719; Euk_kinase.
 DR InterPro; IPR000536; Hormone_rec_1ig.
 DR InterPro; IPR001723; Scdhm_receptor.
 DR InterPro; IPR001245; Tyr_kinase.
 DR InterPro; IPR01628; Znf_Casteroid.
 DR Pfam; PF00104; hormone_rec; 1.
 DR Pfam; PF00069; kinase_rec; 1.
 DR SMART; SM00105; zf-C4; 1.
 DR PRINTS; PR00398; STRDHOMONER.
 DR PRINTS; PR00047; STROIDFINGER.
 DR ProDom; PD000001; Euk_kinase; 1.
 DR ProDom; PD000035; Znf_Casteroid; 1.
 DR SMART; SM00430; HOL1; 1.
 DR SMART; SM00219; TyKc; 1.
 DR SMART; SM00399; ZNF_C4; 1.
 DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
 KW transcription regulation; Transferase; Tyrosine-protein kinase;
 KW Zinc-finger
 SQ SEQUENCE 962 AA; 108320 MW; 3C5ED791E895CE CRC64;
 Query Match 24.1%; Score 1654; DB 15; Length 962;
 Best Local Similarity 51.7%; Pred. No. 3.5e-116;
 Matches 358; Conservative 73; Mismatches 142; Indels 119; Gaps 18;
 QY 547 VEEGRVLOGLPRE YNNAR-HCLP-----CHPEQC 574
 DB 354 IEKQESYLAFFHYIYRKHNIPHEWSKILMKVADLRMIGAYHASFIMKVECPTELS 413
 QY 575 PONGSVTCFGEADQCAVCAHYKDPFCVACPSGVKPDLSYPIKPFDEGACOPCI 634
 DB 414 PQE-----VGP--DHCKCAHFIDGPHCVKACAGVLGENDTL-VKRYADANAVALCQHP 465
 QY 635 NCHSCVLDLDDKGCPCAPQASPLTISAVV-GILLVVLGVVFGILIKRQOKIRKXTM 693
 DB 466 NCTRGCKGPGLEGP--NGSKTPSIAAGVGGILCLVAVGIGIGLYLRR-HIVKRTL 521
 QY 694 RRLQETELVEPLTPSGAMPNOAMRIKETELRKVKVLGSGAFGVYKGMIPGGENVK 753
 DB 522 RRLQEREIVLPLTPSEANNOAHRIKETEFKVKVLFQAFGVYKGLMIPSEKVT 581
 QY 754 IPAIKVLRRENTSPKANKELIDEAYVAVGSPVYSRLIGICLTSTVOLVQIMPYGLL 813
 DB 582 IPAIKELRENTSPKANKELIDEAYVAVSVDPNVCRLIGICLTSTVOLVQIMPYGLL 641
 QY 814 DHVENNGRLSGODLLNMCQIAKMSYLEDVLRDLAARVLYKSNHVKITDFGLA 873
 DB 642 DYIHKONIGSODLLNMCQIAKMSYLEERHVRDLAARVLYKTOHVKITDFGLA 701
 QY 874 RLIDIDETEHADGKVPKIKMALESILRRFTHSDVMSYGVTVWELMTFGAKPYDGI 933
 DB 702 KQCADKEHYHAEGKVPKIKMALESILHRIYTHSDVMSYGVTVWELMTFGSKYDGI 761
 QY 934 AREPDLLEKGERLPQPICTIVYMIYKCMWIDSECPREELVSESNARPORFV 993
 DB 762 ASEISSVLEKGERLPQPICTIVYMIYKCMWSDSPKRELDIAEFSKMARPPRYL 821
 QY 994 VIQ-NEDLGASPLSTFTYSLLEDMDGDLVAEEVLPQOGFCFCDPAAGAGVHHR 1052
 DB 822 VIQDERMHLPSPTDSFRTLMEEEDMEDIADAEVLVPHQGF----- 866

QY 1053 HRSSSTSGGDLTLGLPSEEPAPRPLAPSEGAGVDFDGLGMAKGLQS.PTHDP 111
 DB 867 -NSPT-----SRPLSSLSATSN-----NSATTCIRNGCH-- 898
 QY 1113 SPLQRYSEDPVLPSPSTGVAALTCSPQPEVYNQDPVNPQPSPEEGFLPAPAGAT 117
 DB 899 -----PVREDGFL-----PAPEVYNQ--LMPKKSTAMVQNIYNYISLT 936
 QY 1173 -LERAKTLSPCKNGVNDVAFAGAVNPEYL 1203
 DB 937 AISHLPMDSRYN-----SHSTAVDNPEYL 961
 RESULT 14
 Q85468
 ID Q85468 PRELIMINARY; PRT; 545 AA.
 AC Q85468; 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DR 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Avian Erythroblastosis virus (Ts34) v-erbB gene.
 OS Avian erythroblastosis virus.
 OC Viruses, Retroid viruses, Retroviridae, Avian type C retroviruses.
 OX NCBI_TaxID=11861;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA MEDLINE=88217326; PubMed=2897102;
 RX Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
 RT "Common site of mutation in the erbB gene of avian erythroblastosis
 RT virus mutants that are temperature sensitive for transformation.";
 RL Oncogene Res. 1:265-278(1987).
 DR EMBL; X06943; CA30024.1; -.
 DR HSSP; P11362; 1FGK.
 DR InterPro; IPR000719; Euk_kinase.
 DR InterPro; IPR001245; Tyr_kinase.
 DR Pfam; PF00069; kinase; 1.
 DR ProDom; PD000001; Euk_kinase; 1.
 DR SMART; SM00219; TyKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 545 AA; 60899 MW; 140DCB8CQAF8AF4 CRC64;
 Query Match 24.0%; Score 1645; DB 15; Length 545;
 Best Local Similarity 54.9%; Pred. No. 7.2e-116;
 Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15
 QY 584 GPEADQCAVCAHYKDPFCVACPSGVKPDLSYPIKPFDEGACOPPINCHSCVDL 643
 DB 1 GP--DHCKCAHFIDGPHCVKACPAVLGENDTL-VKRYADANAVALCQHPCTRCCKGP 57
 QY 644 DDKCPAEGRASPPLTISAVV-GILLVVLGVVFGILIKRQOKIRKTYMRLLQETEL 702
 DB 58 GLBSCP--NGSKTPSIAAGVGGILCVVGGIGLYLRR-HIVKRTLRLLQERL 113
 QY 703 VEPILTPSGAMPNOAMRIKETELRKVKVLGSGAFSTYKGMIPDGENVKIPVAIKVLR 762
 DB 114 VEPILTPSGAMPNOAMRIKETEFKVKVLFQAFGVYKGLMIPSEKVTIPVAIKEIR 173
 QY 763 ENTPSPKANKELIDEAYVAVGSPVYSRLIGICLTSTVOLVQIMPYGLLDHVENGR 822
 DB 174 EATSPKANKELIDEAYVAVSVDPNVCRLIGICLTSTVOLVQIMPYGLLDYIREHKON 233
 QY 823 LSGODLLNMCQIAKMSYLEDVLRDLAARVLYKSNHVKITDFGLARLIDETE 882
 DB 234 IGSQVYLLNMCQIAKMSYLEERHVRDLAARVLYKTPQVVKITDFGLAQCADKE 293
 QY 883 YHADGKVPKIKMALESILRRFTHSDVMSYGVTVWELMTGAKPYGICPAREPDLLE 942
 DB 294 YHAEGKVPKIKMALESILHRIYTHSDVMSYGVTVWELMTFGSKPYDGIASEISSVLE 353

```

QY 943 KGERLPPICTIDYVMIMVKCMIDSPRPRELVESESRMARDQRFVIO-NEDLG 1001
DB 354 KGERLPPICTIDYVMIMVKCMIDSPRPRELVESESRMARDQRFVIO-NEDLG 413
QY 1002 PASPLDSTFYRSLLEDDMDLVDAEYLVPOQGFCEPDPAAGGVNHRHRSSTRSG 1061
DB 414 LPSPTDSKRYFTLMEEDEMDIVDAEYLVPHQGF-----NSPT--- 454
QY 1062 GGDLTGLLEPSEEAAPRSPL-----APSEGAGSDVFQDGLMGAAKGLSPTIDPSPLQ 1116
DB 455 -----SRPLUSSLSATSNNAATNIDNKG-----H----- 481
QY 1117 RYSEDPVPLPSETDGYAPLTCSPQPEYVQNPDPVQPPSPREGRPLAARPAAGAT-LER 1175
DB 482 -----PYREDQFL-----PAPRYVNO-LMPKPESTAVVQIVYISLTATSK 523
QY 1176 AKTLSPGKGVKVDVAFARGANVENPEYL 1203
DB 524 LPMDSRYQN-----SHSTAVDNEYL 544

RESULT 15
Q9WVF5 PRELIMINARY: PRT: 655 AA.
AC Q9WVF5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor
   isoform 3).
GN EGFR.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Lampand A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Mahle N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode
   Carboxy-Terminal Truncated Receptors.";
RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTAJ;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampand A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Mahle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
   mouse alternative Egr transcripts encoding truncated receptor
   isoforms.";
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RC MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Kuehl P., Lewis S., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Schriml L.M., Staudt F., Suzuki K., Tomita K., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boileau D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gusninch S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Marzarelli J., Mombaerts P.,
RA Morone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

```

```

RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohno S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AF124513; AAD4149.1; -
DR EMBL; AF275366; AAG28047.1; -
DR EMBL; AF275364; AAG28047.1; JOINED.
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AK004944; BAB23608.1; -
DR EMBL; AK004883; BAB23641.1; -
DR EMBL; AK004911; BAB23662.1; -
DR MGI; 95294; Egr.
DR InterPro; IPR00494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

```

```

Query Match 21.4%; Score 1464.5; DB 11; Length 655;
Best Local Similarity 43.4%; Pred. No. 4.4e-102;
Matches 281; Conservative 98; Mismatches 235; Indels 33; Gaps 10
QY 11 LLLALPREGA--STGYCTGDMTLRLPASPETHLDMLRHLYGCGGVVQGNLETLPTN 68
DB 14 LITLCAAGGLBEKKYCGTGSNFLTQGTFFEDHFLSLQMYNNCEVVLGNLETLTVQRN 73
QY 69 ASLSFLQDIQEVQGVLLAHNOVQVPLQLRLIVRGTLQFEDNYALAVLDNGDPLNNTTP 124
DB 74 YDLSFLKTIQVAGVILALNTVERILENLIQIRGNALYENTATLALISN----- 124
QY 129 VTGASPGRLREQLRLSLETLIKGVLLQRPQLCYQOTILMKDI----FKNNQALATLI 184
DB 125 -YGTNRTELRLPMKNLOETILIGAVRSNPNPICNMOTIQMRDVLQNVFNSMNSMDL--- 186
QY 185 DTNRGRACHPCSPMCKSGRCWSESEDCQSLTRTVCAAGCA-RCKGPLPTDCCGECGCAAG 243
DB 181 -QSHPSGCPKDPBPSCMGSCGGEENQGLTKIKACQSHCRGRGSPSDCCNQCAG 235
QY 244 CTGPAHSQCLACHFNHSIGCELHCPALVTYNTDFESMNPBGRTYFGASCYTACPYNY 303
DB 240 CTGPRESQCLVQCFQDEATCKDCPPLMYPPTYQMDVNPBGKYSFGATCYVKKCPRNY 295
QY 304 LSTVGSCTLVCPPLHNOEYTAEDGTQRCCKSPCARVYGLGM--FNNFTVSFWLRVPK 361
DB 300 VVTDHSGCVRACGDYIEV-EDGIRKCKCKDGPCKRVKVCNGIGIGEKD-TLS----- 350
QY 352 VSASHLEEFACCKKIFSSLAFLPESPDGASNTAPQYIRANSKFIOITELTGYLYISAW 421
DB 351 INATIKFKFKYCTAISGLHLPLVAFKDSFTRPPLDPRELETLKVKKEITGFLIIQAW 410
QY 422 PDSLPDSVFPNLQVINGRILHNAVYLLTQGISLIGRLSRELCSGLALHNHTHIC 481
DB 411 PDNWTDLAFENLILNGRTKQHQFSLAVAGLNTISLGRSLKEISDGVVITSGNRNIC 470
QY 482 FVHTVPMQDLFRNPHQALLHTANPEDECEVGEGLACHQLCARGHGMPGPPTQCVNCSQFL 541
DB 471 YANTINWKKLFGTNGTKTKIMNRAEKDCAVNVHVCNPLCSSGCMGPEPRDVCQNNIS 530
QY 542 RGQCEVECEKTLQGLPREYVNAARCLFCHPECCQPNQSVTCGPEADQCVACAHYQDPP 601
DB 531 RGRCEVEKCNLTLEGEPEFVENSECICQHECJFOANNITCTGRGPDNCTQCAHYIDGFH 590
QY 602 CVACPSGVKPDLSYMPINKFPDEGACQCPINCTFSCTVDLDDKCG 648
DB 591 CVKICPAGIMENNTL-VMKYADANVCHLCHANCYTGCGAGPLQCG 636

```

Search completed: July 22, 2003, 09:01:17
 Job time : 54.5887 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 38.1589 Seconds

(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-369-383-14

Perfect score: 6853

Sequence: 1 MELALCRWGLLALPPGA.....TFKGTPTAENPEYLGLDVP 1261

Scoring table: BLOSUM62

Gap: 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	6661	97.2	1255	21	AA92620 Human heregulin 2
2	6661	97.2	1255	22	AAE12130 Human tyrosine kin
3	6661	97.2	1255	22	AA860167 Human transgene pla
4	6661	97.2	1255	23	AA174545 Human HER2 (ErbB2)
5	6655	97.1	1255	17	AA011111 Human HER-2/neu protei
6	6655	97.1	1255	20	AA92406 Human HER-2/neu on
7	6655	97.1	1255	21	AA821198 Human HER-2/neu pr
8	6655	97.1	1255	21	AA84780 Amino acid sequenc
9	6655	97.1	1255	22	AA85458 Human HER-2/neu pr
10	6655	97.1	1255	22	AA88267 Human HER-2/neu amino aci

11	6655	97.1	1255	23	AAE24067 Human Her-2 protei
12	6655	97.1	1255	23	AAE20479 Human Her-2/neu pr
13	6655	97.1	1255	23	AA051143 Human Her-2/neu on
14	6655	97.1	1255	23	AA177114 Human Her-2/neu pr
15	6616	96.5	1433	14	AA39568 Sequence of c-erbB
16	6491	94.7	1223	23	AA98923 Human breast cance
17	6338	92.5	1200	21	AA821208 Human HER-2/neu pr
18	5863.5	85.6	1256	23	AA821199 Rat HER-2/neu prot
19	5863.5	85.6	1256	23	AA821199 Rat Her-2/neu onc
20	5842.5	85.3	1256	22	AA821206 Mouse Her-2/neu pr
21	5842.5	85.3	1256	22	AA821206 Amino acid sequenc
22	5842.5	85.3	1256	23	AA821203 Mouse Her-2/neu or
23	4741	69.2	919	23	AA821203 Human HER-2/neu fu
24	4741	69.2	919	23	AA821203 Her-2/neu extracel
25	3992.5	58.3	926	23	AA821203 Mouse Her-2/neu ex
26	3992.5	58.3	926	23	AA821203 Mouse Her-2/neu ex
27	3625	52.9	712	21	AA821204 Human HER-2/neu ex
28	3625	52.9	712	21	AA821204 Her-2/neu extracel
29	3479	50.8	782	18	AA821200 Her-2-GM-CSF immunc
30	3477	50.7	653	21	AA821200 Extracellular HER-
31	3477	50.7	653	23	AA821200 Human Her-2/neu or
32	3439	50.2	645	22	AA860408 Human ErbB2 oncopr
33	3439	50.2	645	22	AA860408 Human ErbB2 extrac
34	3374	49.2	951	21	AA821203 DC86FV-erbB2BC fu
35	3271	47.7	624	11	AA821203 Human EGF receptor
36	3101	45.3	1210	21	AA819259 Human Her-1 protei
37	3101	45.3	1210	23	AA823019 Human epidermal gr
38	3101	45.3	1210	23	AA823019 Amino acid sequenc
39	3101	45.3	1210	23	AA823019 Human EGF receptor
40	3099	45.2	1210	22	AA868420 Human Her-1 protei
41	3084	45.0	583	23	AA820483 Amino acid sequenc
42	3084	45.0	587	23	AA820481 Human protein for
43	3083	45.0	589	23	AA820482 Human protein for
44	3083	45.0	600	23	AA820482 Human protein for
45	3060	44.7	1210	23	AB851768 Human epidermal gr

ALIGNMENTS

```
RESULT 1
ID AA92620
AA92620 standard; Protein; 1255 AA.
AC AA92620;
XX 10-AUG-2000 (first entry)
XX Human heregulin 2 (Her2).
DE Human heregulin 2 (Her2).
XX Her2; vaccination; cytotoxic T-lymphocyte immunity;
XX self-protein; cancer; breast cancer; prostate cancer;
XX cell-associated peptide antigen; foreign epitope.
XX Homo sapiens.
XX
XX Key
XX Domain
XX Location/Qualifiers
FT 1..173 /label= "N-terminal"
FT /note= "mature polypeptide"
FT 5..25 /label= "insertion_region"
FT /note= "suitable for foreign epitope insertion"
FT 59..73 /label= "insertion_region"
FT /note= "suitable for foreign epitope insertion"
FT 103..117 /label= "insertion_region"
FT /note= "suitable for foreign epitope insertion"
FT 149..163 /label= "insertion_region"
FT /note= "suitable for foreign epitope insertion"
FT 174..323 /label= "insertion"
FT Domain
```

FT /label= Cysteine_rich_domain
 FT 210..224
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 250..264
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 324..483
 FT /label= ligand_binding_domain
 FT 325..339
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 359..383
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 465..479
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 484..623
 FT /label= Cysteine_rich_domain
 FT 579..593
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 624..654
 FT /label= Transmembrane_domain
 FT 632..652
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 653..667
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 655..1010
 FT /label= Tyrosine_kinase_domain
 FT 661..675
 FT /label= insertion_region
 FT /label= "suitable for foreign epitope insertion"
 FT Region
 FT 695..709
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 710..730
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 1011..1235
 FT /label= C-terminal_domain
 PN MO200020027-A2.
 PD 13-APR-2000.
 PF 05-OCT-1999; 99WO-DK00525.
 PR 05-OCT-1998; 98DK-0001261.
 PR 20-OCT-1998; 98US-0105011.
 XX (MEB1-) M & E BIOTECH AS.
 PA Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
 PI Gautam A, Birk P, Karlsson G;
 XX MPI: 2000-349917/30.
 DR N-PSDB; AAA09455.
 PT Inducing immune responses to weakly immunogenic, tumor associated
 PT peptide antigens for the treatment of breast and prostate cancer
 XX
 XX Claim 62; Page 193-198; 220pp; English.
 CC This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 CC Her2 can be used in the claimed method as an autovaccine to induce a CTL
 CC response. Subdominant CTL epitopes, antibody binding regions and
 CC cysteine residues involved in disulfide bonds are preserved in the
 CC immunogenized forms. Regions suitable for the insertion of foreign T
 CC helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the animals immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respectiv
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/breast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.
 CC XX

SO Sequence 1255 AA;

Query Match 97.2%; Score 6661; DB 21; Length 1255;
 Best Local Similarity 97.5%; Pred. No. 0;
 Matches 1231; Conservative 10; Mismatches 13; Indels 8; Gaps

QY 1 MELALCRWGLLALPPGASTQVCTGDMKALPASPEHMLMHLVQCCVVGSL 60
 DB 1 MELALCRWGLLALPPGASTQVCTGDMKALPASPEHMLMHLVQCCVVGSL 60
 QY 61 ELTYLPTNASLFLQDIQEVGYLIANQVQVPLQRLIRVGTQFEDNYALAVDNG 120
 DB 61 ELTYLPTNASLFLQDIQEVGYLIANQVQVPLQRLIRVGTQFEDNYALAVDNG 120
 QY 121 DPLNNTTPTVGTASPGGLBELRLSLTELKKGVLIOGNPOLCYDTTLKMDIFKHNQLA 180
 DB 121 DPLNNTTPTVGTASPGGLBELRLSLTELKKGVLIOGNPOLCYDTTLKMDIFKHNQLA 180
 QY 181 LTLIDTNSRACHPSPCKSKSRMGSSSEDCGLRTVACGAGCARCKGPLPDDCCEQC 240
 DB 181 LTLIDTNSRACHPSPCKSKSRMGSSSEDCGLRTVACGAGCARCKGPLPDDCCEQC 240
 QY 241 AAGCTGPRHSDCLACLHNHSGICELHCALVTYNTDFFSMPNPEGRYTGCACVTACP 300
 DB 241 AAGCTGPRHSDCLACLHNHSGICELHCALVTYNTDFFSMPNPEGRYTGCACVTACP 300
 QY 301 YNYLSTDVGSCTVCPILHNOVTEDEGTQRCCKSKCAVCVGLGFNNFTYSFWLR-V 355
 DB 301 YNYLSTDVGSCTVCPILHNOVTEDEGTQRCCKSKCAVCVGLGFNNFTYSFWLR-V 355
 QY 360 PKVASHLBEPAGCKKIFGSLAFPLPESFGDGPASNTAPQYIKANSKFICTELTGYLYIS 419
 DB 360 PKVASHLBEPAGCKKIFGSLAFPLPESFGDGPASNTAPQYIKANSKFICTELTGYLYIS 419
 QY 420 AMPDGLPPLSVFQNLQVYRGRILNNGAYSLTLQGLGTSWGLRLSGGLALIHNNTH 479
 DB 420 AMPDGLPPLSVFQNLQVYRGRILNNGAYSLTLQGLGTSWGLRLSGGLALIHNNTH 479
 QY 479 LCFVHTVPMQDLFNPHQALHTANRDEDECVSGELACHQLCARHGCMGPPIQVNCNQ 533
 DB 479 LCFVHTVPMQDLFNPHQALHTANRDEDECVSGELACHQLCARHGCMGPPIQVNCNQ 533
 QY 540 FLRQGEVCEBGRVLIQIIPREYVNAHCLPCHPSCOPONGSVTCFGBADQCVACAHYKP 599
 DB 540 FLRQGEVCEBGRVLIQIIPREYVNAHCLPCHPSCOPONGSVTCFGBADQCVACAHYKP 599
 QY 599 PFCVAPCSGKVPPLSMPIWKPDEDEGACOPINCTHSCVLDKGCAPAEGRASPLTS 659
 DB 599 PFCVAPCSGKVPPLSMPIWKPDEDEGACOPINCTHSCVLDKGCAPAEGRASPLTS 659
 QY 660 IVSAVGLLVVLVGVFGLIKRRQOKIRKTYMRRLQETELVEPLTPSGAMPNQAQMR 719
 DB 660 IVSAVGLLVVLVGVFGLIKRRQOKIRKTYMRRLQETELVEPLTPSGAMPNQAQMR 719
 QY 720 ILKTELIRKXKVVGSGAGFVYVGIWIPQENKIKYVAIKVLRNTSPKANKILDEAVY 779
 DB 720 ILKTELIRKXKVVGSGAGFVYVGIWIPQENKIKYVAIKVLRNTSPKANKILDEAVY 779
 QY 779 ILKTELIRKXKVVGSGAGFVYVGIWIPQENKIKYVAIKVLRNTSPKANKILDEAVY 779
 DB 779 ILKTELIRKXKVVGSGAGFVYVGIWIPQENKIKYVAIKVLRNTSPKANKILDEAVY 779

QY	786	MAGVSPVYSLRLGLCTLSTVLTQMLPQYGLLHVHNRNRLCSODLLNWCQIACM	839
Db	774	MAGVSPVYSLRLGLCTLSTVLTQMLPQYGLLHVHNRNRLCSODLLNWCQIACM	833
QY	840	SYLEEVRLVHRLDAAARNVLYKSPNNVYKLTDFGLARLIDIDEYHADGAKVPIKMMALIS	899
Db	834	SYLEEVRLVHRLDAAARNVLYKSPNNVYKLTDFGLARLIDIDEYHADGAKVPIKMMALIS	893
QY	900	ILRRRFTTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPEICTTIDVYM	959
Db	894	ILRRRFTTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPEICTTIDVYM	953
QY	960	INVKCMIMIDSECRPPFRRELVSSEFSMARDDPQRFVYTIQWEDLGAPASPLDSTFRRILLEDD	1019
Db	954	INVKCMIMIDSECRPPFRRELVSSEFSMARDDPQRFVYTIQWEDLGAPASPLDSTFRRILLEDD	1013
QY	1020	MGDLVDAEYELVPQCGFFCPCPDPAFAGGMVHRRHSSSTRSGGDLTLGLEPSEEARPS	1079
Db	1014	MGDLVDAEYELVPQCGFFCPCPDPAFAGGMVHRRHSSSTRSGGDLTLGLEPSEEARPS	1073
QY	1080	PLABEAGASDVFDGDLGMGAAKGIQSLPTHDPSLQRYSEDPTVPLPSETDGYVAPLTC	1139
Db	1074	PLABEAGASDVFDGDLGMGAAKGIQSLPTHDPSLQRYSEDPTVPLPSETDGYVAPLTC	1133
QY	1140	SPQPEYVNOQDVRRPQPPSPRESGPLPAAPAGATLERAKTILSPGKGVWYKDYPAFPGAVEN	1199
Db	1134	SPQPEYVNOQDVRRPQPPSPRESGPLPAAPAGATLERAKTILSPGKGVWYKDYPAFPGAVEN	1193
QY	1200	PEYLTTPQGGAPQPPHPPPAFSPAFENLYWDDPPERGAPSTFKGPTAENPEYLGIDV	1259
Db	1194	PEYLTTPQGGAPQPPHPPPAFSPAFENLYWDDPPERGAPSTFKGPTAENPEYLGIDV	1253
QY	1260	PV 1261	
Db	1254	PV 1255	
RESULT 2			
ID	AAE12130	standard; Protein; 1255 AA.	
XX	AAE12130;		
AC			
DT	18-DEC-2001	(first entry)	
XX			
DE	Human tyrosine kinase-type receptor, HER-2.		
KM	Therapeutic compound; major histocompatibility complex; vaccine;		
KM	antigenic peptide; MHC; immunoregulatory; immune response; HER-2;		
KM	adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;		
KM	antigen presenting cell; human; tyrosine kinase-type receptor.		
OS	Homo sapiens.		
XX			
HH	Key	Location/Qualifiers	
FT	Region	774..782	
XX		note= "Antigenic epitope"	
PN	WC0200168677-A2.		
XX			
PD	20-SEP-2001.		
XX			
PF	16-MAR-2001; 2001WO-US40328.		
XX			
PR	16-MAR-2000; 2000US-0527487.		
XX			
PA	(GENZ) GENZYME CORP.		
XX			
FI	Nicolete CA;		
XX			
WP	WPI; 2001-616284/71.		
XX	OR		
XX	N-PSDB; AAD19731.		
XX			

PT	Novel synthetic therapeutic compound for inducing immune response and
PT	for use in adoptive immunotherapy, has enhanced binding to major
PT	histocompatibility molecules and enhanced immunoregulatory properties
XX	-
PS	Claim 4; Page 63-67; 69pp; English.
XX	
CC	The invention relates to synthetic therapeutic compounds (antigenic
CC	peptides) with enhanced binding to major histocompatibility complex
CC	(MHC) molecules and enhanced immunoregulatory properties relative
CC	to their natural counterparts. Compounds of the invention are useful
CC	for inducing an immune response in a subject and for use in adoptive
CC	immunotherapy. They are useful as components of anti-cancer vaccines
CC	and to expand immune effector cells that are specific for cancers
CC	characterised by expression of the breast cancer antigen, HER-2.
CC	Polynucleotides that encode peptides of the invention are useful as
CC	hybridisation probes and as primers for the detection of genes of gene
CC	transcripts that are expressed in antigen presenting cells (APCs), to
CC	confirm transduction of polynucleotides into host cells. The present
CC	sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC	of the invention are designed based on the HER-2 antigenic peptide
CC	(774-782).
XX	
SO	Sequence 1255 AA;
Query Match	97.2%; Score 6661; DB 22; Length 1255;
Best Local Similarity	97.5%; Pred. No. 0;
Matches 1231; Conservative 10; Mismatches 13; Indels 8; Gaps	
QY	1 MELAALCRWGLLALLPFGAASQVCTGTDMLRLPASSETHLDMRLHYOCQVVGNTL 60
DB	1 MELALCRWGLLALLPFGAASQVCTGTDMLRLPASSETHLDMRLHYOCQVVGNTL 60
QY	61 ELTYLPFNASLSFLQDIOEQGVYLLIAHQVQVPLQRLIRYRGQLFEDNALAVLNG 12
DB	61 ELTYLPFNASLSFLQDIOEQGVYLLIAHQVQVPLQRLIRYRGQLFEDNALAVLNG 12
QY	121 DPLNNTTPVTGASPGGLREIQLRSLTEILKGGVLIQRNPQLCYODTILMKDIFKNNOLA 180
DB	121 DPLNNTTPVTGASPGGLREIQLRSLTEILKGGVLIQRNPQLCYODTILMKDIFKNNOLA 180
QY	181 LTLIDNRSRACHPCSPCKGSGCWSESEDDOSILTRTYCAGGACRCKGPIPTDCHEQC 24
DB	181 LTLIDNRSRACHPCSPCKGSGCWSESEDDQSILTRYCAGGACRCKGPIPTDCHEQC 24
QY	241 AAGCTGKXSDCLACILHFNHSGICELHCPALTYNTDTFESYPNRGRYTFGASCVTACP 300
DB	241 AAGCTGKXSDCLACILHFNHSGICELHCPALTYNTDTFESYPNRGRYTFGASCVTACP 300
QY	301 YNYLSTDVGSCTLVCLHNOEVTAEEDGTORCKGSKPCARVVCYGLGMFNNTVGSFWLR-V 355
DB	301 YNYLSTDVGSCTLVCLHNOEVTAEEDGTORCKGSKPCARVVCYGLGMFNNTVGSFWLR-V 355
QY	360 PKYSASHLEEFACCKKIFGSLAFLESPFGDPASTAPQVYKANSKFGITELGYLIS 415
DB	360 PKYSASHLEEFACCKKIFGSLAFLESPFGDPASTAPQVYKANSKFGITELGYLIS 415
QY	420 AMPDSDPDSVFQNLQVIRGRILHNGAVSLTQGLIGISWLGIRSLRELGSLALIHNTH 479
DB	420 AMPDSDPDSVFQNLQVIRGRILHNGAVSLTQGLIGISWLGIRSLRELGSLALIHNTH 479
QY	480 LCEVNHVPMQDLFRNPHQALHTNARPEDECVGSLAHCILCAGHCHGPRPTCCVNCQ 539
DB	480 LCEVNHVPMQDLFRNPHQALHTNARPEDECVGSLAHCILCAGHCHGPRPTCCVNCQ 539
QY	540 FLRGOECVEECRYLQGLPREYVNAHNCILPCHEQCQPONGSVTFCGPEADQCVAICAHYKDP 599
DB	540 FLRGOECVEECRYLQGLPREYVNAHNCILPCHEQCQPONGSVTFCGPEADQCVAICAHYKDP 599
QY	594 PFCVACSPGVKVDLSYMPIMKFPDEEACACPCPLNCHSGCVLDIDDGCAEAGASLTLS 653
DB	594 PFCVACSPGVKVDLSYMPIMKFPDEEACACPCPLNCHSGCVLDIDDGCAEAGASLTLS 653

```

QY 660 IVSAVVGLLVVVLGVVFGILLIKRQOKIRKXTMRLLQETELVEPLTPSGAMPNOAQR 719
DB 654 IVSAVVGLLVVVLGVVFGILLIKRQOKIRKXTMRLLQETELVEPLTPSGAMPNOAQR 713
QY 720 ILKETELRKVKVLGSAFGTVKGIWIPGENVKIPVAIKVLRNTSPKANKELIDEAYV 779
DB 714 ILKETELRKVKVLGSAFGTVKGIWIPGENVKIPVAIKVLRNTSPKANKELIDEAYV 773
QY 780 MAGVGSPIVSRLLIGLITSTVQLVQLMPYGLLDHVENGRGLSQQLLWMCQIAKGM 839
DB 774 MAGVGSPIVSRLLIGLITSTVQLVQLMPYGLLDHVENGRGLSQQLLWMCQIAKGM 833
QY 840 SYLEDVRLVHRDLAARNVUKSPNHVKITDPGLAFLDIDETEHADGKVPKIMWALE 899
DB 834 SYLEDVRLVHRDLAARNVUKSPNHVKITDPGLAFLDIDETEHADGKVPKIMWALE 893
QY 900 ILRRRFTHOSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPPOPICTIVYM 959
DB 894 ILRRRFTHOSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPPOPICTIVYM 953
QY 960 IMVKCMMIDSECRPRPRELVSEFSMMAADPORFVVIQNEDLGPASPDLSTFYRLDDDD 1019
DB 954 IMVKCMMIDSECRPRPRELVSEFSMMAADPORFVVIQNEDLGPASPDLSTFYRLDDDD 1013
QY 1020 MGDLYDAEEYLVPOOGFPCPDPAFGAGMWHHRSSSTRSGGDLTLGLEPSEEAPRS 1079
DB 1014 MGDLYDAEEYLVPOOGFPCPDPAFGAGMWHHRSSSTRSGGDLTLGLEPSEEAPRS 1073
QY 1080 PLAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLCRYSEDPTVPLPSETDGVVAALTC 1139
DB 1074 PLAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLCRYSEDPTVPLPSETDGVVAALTC 1133
QY 1140 SPQPEYVNOPDVPRQPPSPREGPLPAARPAGATLERATLSFGKXGVKDYFAFGAVEN 1199
DB 1134 SPQPEYVNOPDVPRQPPSPREGPLPAARPAGATLERATLSFGKXGVKDYFAFGAVEN 1193
QY 1200 PEYLTPQGGAAQPHPHPAFSPAFDNLTYMDQDPPERGAPSTFKGTPTAENPEYLGLDV 1259
DB 1194 PEYLTPQGGAAQPHPHPAFSPAFDNLTYMDQDPPERGAPSTFKGTPTAENPEYLGLDV 1253
QY 1260 PV 1261
DB 1254 PV 1255

RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
XX
AC AAB60167;
XX
DT 03-APR-2001 (first entry)
XX
DE HER2 transgene plasmid construct encoded protein.
XX
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
XX
OS antibody.
XX
OS Homo sapiens.
XX
OS Synthetic.
XX
PN WO200100244-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000MO-US17229.
XX
PR 25-JUN-1999; 99US-0141316
XX
PR 16-MAR-2000; 2000US-0189844.
XX
PA (GETH ) GENENTECH INC.
XX
PI Erickson S, Schwall R;

```

```

XX
DR MPI: 2001-061962/07.
DB N-PSDB; AAF24297.
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid -
XX
PS Example 3; Fig 4; 92pp; English.
XX
CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;
Query Match 97.2%; Score 6661; DB 22; Length 1255;
Best Local Similarity 97.5%; Pred. No. 0;
Matches 1231; Conservative 10; Mismatches 13; Indels 8; Gaps :
QY 1 MELALCRWGLLALPPGAASTQVCTGTDMKRLPASPEHLDMLHLYOGCQVVOGNL 60
DB 1 MELALCRWGLLALPPGAASTQVCTGTDMKRLPASPEHLDMLHLYOGCQVVOGNL 60
QY 61 ELTYVPTNASLFIODIOEVQGYVLIANOVQVPLQRLVRGTQLPEDNYALAVLDNG 124
DB 61 ELTYVPTNASLFIODIOEVQGYVLIANOVQVPLQRLVRGTQLPEDNYALAVLDNG 124
QY 121 DELNNTTPTVGTASPGRLRELOLRSLTELKQGVLIQBNPOLCYQDITLMKIDFHKNNQLA 184
DB 121 DELNNTTPTVGTASPGRLRELOLRSLTELKQGVLIQBNPOLCYQDITLMKIDFHKNNQLA 184
QY 181 LTLIDTNSRBACHPECPMKCKSRMGSSSEDCSLRTVACAGGARCKGFLPTDCHEOC 244
DB 181 LTLIDTNSRBACHPECPMKCKSRMGSSSEDCSLRTVACAGGARCKGFLPTDCHEOC 244
QY 241 AAGCTGPHGSDCLACLNHNSGICELHCPALVTYNTDFTESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPHGSDCLACLNHNSGICELHCPALVTYNTDFTESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPLHNOVETLEDGTQRECKSKPCARVCYGLGNFNFTVSPWLR-V 355
DB 301 YNYLSTDVGSCTLVCPLHNOVETLEDGTQRECKSKPCARVCYGLGNFNFTVSPWLR-V 355
QY 360 PKVASHLSEFPAGCKITFGSLAFPESEFDGPPASNTAPQYIKANSKFIGITELTGYLYIS 415
DB 360 PKVASHLSEFPAGCKITFGSLAFPESEFDGPPASNTAPQYIKANSKFIGITELTGYLYIS 415
QY 420 AMPDSLPLSVFQNLQVIRGRILNNGAYSLTLQGLISWLGRLSLRELSGGLALIHNNTH 479
DB 420 AMPDSLPLSVFQNLQVIRGRILNNGAYSLTLQGLISWLGRLSLRELSGGLALIHNNTH 479
QY 480 LCFYHTVPMQDLFENPHQALHTNARPEDEOVGGGLACHOLCARGHMGWGPQPCVNCQSG 539
DB 480 LCFYHTVPMQDLFENPHQALHTNARPEDEOVGGGLACHOLCARGHMGWGPQPCVNCQSG 539
QY 540 FLRQGEVCECRVLQGLPREVYVARHCLPCHPEQOPNGSVTCFGPEADQCVACAHYKXP 599
DB 540 FLRQGEVCECRVLQGLPREVYVARHCLPCHPEQOPNGSVTCFGPEADQCVACAHYKXP 599
QY 594 PFCVARGSGVCKPPLSTWPIWKPFDEBGAQCPICNTHSCVLLDXKGCFAEBRASPLTS 653
DB 594 PFCVARGSGVCKPPLSTWPIWKPFDEBGAQCPICNTHSCVLLDXKGCFAEBRASPLTS 653
QY 660 IVSAVVGLLVVVLGVVFGILLIKRQOKIRKXTMRLLQETELVEPLTPSGAMPNOAQR 719
DB 654 IVSAVVGLLVVVLGVVFGILLIKRQOKIRKXTMRLLQETELVEPLTPSGAMPNOAQR 713
QY 720 ILKETELRKVKVLGSAFGTVKGIWIPGENVKIPVAIKVLRNTSPKANKELIDEAYV 779

```

Db 714 ILKETLRKVKVLSGSAFGTVYKGIWPDGENVKIPVALILERENTSPANKELLDEAVV 773
 Qy 780 MAGVSPYVSRLLGICLTSTVQVLTQMPYGCILDHYRENGSLGSDLLNMCMQIAKGM 839
 Db 774 MAGVSPYVSRLLGICLTSTVQVLTQMPYGCILDHYRENGSLGSDLLNMCMQIAKGM 833
 Qy 840 SYLEDVLYHRDLAARVLYKSPNHVKITDFGLARLLDIDETEHADGKVPKIMALES 899
 Db 834 SYLEDVLYHRDLAARVLYKSPNHVKITDFGLARLLDIDETEHADGKVPKIMALES 893
 Qy 900 ILRRFTHSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLQOPICITIDVW 959
 Db 894 ILRRFTHSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLQOPICITIDVW 953
 Qy 960 IMVKCMWIDSECRPRFELVSESRMARDPQRFVVIQNEEDLGASPLDSTFYRSLLEDD 1019
 Db 954 IMVKCMWIDSECRPRFELVSESRMARDPQRFVVIQNEEDLGASPLDSTFYRSLLEDD 1013
 Qy 1020 MGDLYDAEEYLVPOQGFCDPAPGAGGWHHRHSSSTRSGGDLTLGLESEEPAPS 1079
 Db 1014 MGDLYDAEEYLVPOQGFCDPAPGAGGWHHRHSSSTRSGGDLTLGLESEEPAPS 1073
 Qy 1080 PLASEGAGSDVFDGDLGMAKAGLSLPTHPDSPLOQYSEDETVPLPSEITGYVAPLTC 1139
 Db 1074 PLASEGAGSDVFDGDLGMAKAGLSLPTHPDSPLOQYSEDETVPLPSEITGYVAPLTC 1133
 Qy 1140 SPOEYVNOQDVAPQPSREGEPLPAARPAATLEBAKTLSPKNGVQKVAFGAVEN 1199
 Db 1134 SPOEYVNOQDVAPQPSREGEPLPAARPAATLEBAKTLSPKNGVQKVAFGAVEN 1193
 Qy 1200 PEYLTPOGGAPOHPHPAPSPAFDNLVYWDOPPEPRGAPSTFKGTPTAENPEYLGIDV 1259
 Db 1194 PEYLTPOGGAPOHPHPAPSPAFDNLVYWDOPPEPRGAPSTFKGTPTAENPEYLGIDV 1253
 Qy 1260 PV 1261
 Db 1254 PV 1255

RESULT 4
 AAU74545
 ID AAU74545 standard; Protein; 1255 AA.
 AC AAU74545;
 XX
 DT 23-Apr-2002 (first entry)
 XX
 DE Human HER2 (ErbB2) polypeptide.
 XX
 XX Human; HER2; ErbB; epidermal growth factor receptor; receptor;
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
 KW glial disorder; astrocytal disorder; hypothalamic disorder;
 KW glandular disorder; macrophagal disorder; epithelial disorder;
 KW stromal disorder; blastocoelec disorder; inflammatory disorder;
 KW angiogenic disorder; immunological disorder.
 XX
 OS Homo sapiens.
 XX
 PN US2002001587-A1.
 XX
 PD 03-JAN-2002.
 XX
 PF 16-MAR-2001; 2001US-0811123.
 XX
 PR 16-MAR-2000; 2000US-189844P.
 PA 05-OCT-2000; 2000US-238327P.
 XX
 PA (ERIC/) ERICKSON S.
 PA (SCHW/) SCHWALL R.
 PA (SLIW/) SLIWKOWSKI M.
 XX

PI Erickson S, Schwall R, Sliwkowski M;
 XX
 DR WPI: 2002-163686/21.
 DR N-PDB; ABK14058.
 XX
 PT Treating tumour characterised by overexpression of epidermal growth
 PT factor receptor, ErbB or cancer in mammal, comprises admistering
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal -
 XX
 PS Example 3; Fig 7; 93pp; English.
 XX
 CC The invention relates to treating a tumour in a mammal, where the tumour
 CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (ErbB) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses ErbB2. The method is also useful for treating
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
 CC epithelial, stromal, blastocoelec, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)
 CC polypeptide of the invention.
 XX
 SQ Sequence 1255 AA;
 Query March 97.2%; Score 6661; DB 23; Length 1255;
 Best Local Similarity 97.5%; Pred. No. 0;
 Matches 1231; Conservative 10; Mismatches 13; Indels 8; Gaps 2

Qy 1 MELAALCRWGLLALPLPGAASTQVCTGDMKRLRPASPTHLDMLRHYQCQVVGNTL 60
 Db 1 MELAALCRWGLLALPLPGAASTQVCTGDMKRLRPASPTHLDMLRHYQCQVVGNTL 60
 Qy 61 ELTYLPNASTLFDIOIVQGVVLIHANQVQVPLQRIYRIGTQFEDNYALAVLDNG 120
 Db 61 ELTYLPNASTLFDIOIVQGVVLIHANQVQVPLQRIYRIGTQFEDNYALAVLDNG 120
 Qy 121 DPLNNTPTVYTGASPGGLRELOLSLTELKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 180
 Db 121 DPLNNTPTVYTGASPGGLRELOLSLTELKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 180
 Qy 121 DPLNNTPTVYTGASPGGLRELOLSLTELKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 180
 Db 121 DPLNNTPTVYTGASPGGLRELOLSLTELKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 180
 Qy 181 LTLIDNRSBACHPCSPMKSGSRCWGESSEDCQSLTRYCAGGACARCKPPTDCHEQC 240
 Db 181 LTLIDNRSBACHPCSPMKSGSRCWGESSEDCQSLTRYCAGGACARCKPPTDCHEQC 240
 Qy 181 LTLIDNRSBACHPCSPMKSGSRCWGESSEDCQSLTRYCAGGACARCKPPTDCHEQC 240
 Db 181 LTLIDNRSBACHPCSPMKSGSRCWGESSEDCQSLTRYCAGGACARCKPPTDCHEQC 240
 Qy 241 AAGCTGPKHSDCLACHFNHSGICELHCAVLTNTDPTESMPNEGRYTFASCVTACP 300
 Db 241 AAGCTGPKHSDCLACHFNHSGICELHCAVLTNTDPTESMPNEGRYTFASCVTACP 300
 Qy 301 YNYLSTDVSGCTLVCELIHNOEYTAEDGTQRCCKSKPCARVCGYGLGNFTVSWFLR-V 359
 Db 301 YNYLSTDVSGCTLVCELIHNOEYTAEDGTQRCCKSKPCARVCGYGLGNFTVSWFLR-V 359
 Qy 301 YNYLSTDVSGCTLVCELIHNOEYTAEDGTQRCCKSKPCARVCGYGLGNFTVSWFLR-V 359
 Db 301 YNYLSTDVSGCTLVCELIHNOEYTAEDGTQRCCKSKPCARVCGYGLGNFTVSWFLR-V 359
 Qy 360 PKYSASHLEFPAGCKKIFPSLAFLEPSEFGDPASNTAPQYIKANSKFIIGITELGYLIS 419
 Db 360 PKYSASHLEFPAGCKKIFPSLAFLEPSEFGDPASNTAPQYIKANSKFIIGITELGYLIS 419
 Qy 360 PKYSASHLEFPAGCKKIFPSLAFLEPSEFGDPASNTAPQYIKANSKFIIGITELGYLIS 419
 Db 360 PKYSASHLEFPAGCKKIFPSLAFLEPSEFGDPASNTAPQYIKANSKFIIGITELGYLIS 419
 Qy 354 RAVTSANIOEPAGCKKIFPSLAFLEPSEFGDPASNTAPQYIKANSKFIIGITELGYLIS 413
 Db 354 RAVTSANIOEPAGCKKIFPSLAFLEPSEFGDPASNTAPQYIKANSKFIIGITELGYLIS 413
 Qy 420 AMPDSLPLDSVFQNLQVIRGILIHNGAVSLTQGLISMLGRLSRLREGSLALIHNTH 479
 Db 420 AMPDSLPLDSVFQNLQVIRGILIHNGAVSLTQGLISMLGRLSRLREGSLALIHNTH 479
 Qy 414 AMPDSLPLDSVFQNLQVIRGILIHNGAVSLTQGLISMLGRLSRLREGSLALIHNTH 473
 Db 414 AMPDSLPLDSVFQNLQVIRGILIHNGAVSLTQGLISMLGRLSRLREGSLALIHNTH 473
 Qy 480 LCFVHTVPMDOULFRNPHOALLTANRPEDECVGEGSLACHQLCARGHCGPPTQCVNCSQ 539
 Db 480 LCFVHTVPMDOULFRNPHOALLTANRPEDECVGEGSLACHQLCARGHCGPPTQCVNCSQ 539
 Qy 474 LCFVHTVPMDOULFRNPHOALLTANRPEDECVGEGSLACHQLCARGHCGPPTQCVNCSQ 533
 Db 474 LCFVHTVPMDOULFRNPHOALLTANRPEDECVGEGSLACHQLCARGHCGPPTQCVNCSQ 533
 Qy 540 FLRGQECVEGCVLQGLPREYVNAHNCPCPECCOPQNGSVTCGPEADQCVAAAHKXP 599
 Db 540 FLRGQECVEGCVLQGLPREYVNAHNCPCPECCOPQNGSVTCGPEADQCVAAAHKXP 599
 Qy 534 FLRGQECVEGCVLQGLPREYVNAHNCPCPECCOPQNGSVTCGPEADQCVAAAHKXP 593
 Db 534 FLRGQECVEGCVLQGLPREYVNAHNCPCPECCOPQNGSVTCGPEADQCVAAAHKXP 593
 Qy 600 PFCVACPSGVKPDLSYMPIMKFPDEGACQPCPINCTHSCVDLDDKQCAEGRASPLTS 659
 Db 600 PFCVACPSGVKPDLSYMPIMKFPDEGACQPCPINCTHSCVDLDDKQCAEGRASPLTS 659

```

Db      594 PFCVACRCSGVFDLSYMPIMKFPDEBACQPCCLINCHSVDLDDGCCAERASLTLS 653
Qy      660 TVSAVAVGLLVVTVGVFGILIKRQOKIRKTYTRRLLOETELVEPLTPSGAMPNOQMR 719
Db      654 TVSAVAVGLLVVTVGVFGILIKRQOKIRKTYTRRLLOETELVEPLTPSGAMPNOQMR 713
Qy      720 ILKETELRKVKVLSGAFGVYKGIWIPDGENVYKIPVAIKVLRNTPSPKANKELDEAVY 779
Db      714 ILKETELRKVKVLSGAFGVYKGIWIPDGENVYKIPVAIKVLRNTPSPKANKELDEAVY 773
Qy      780 MAGVSPVYSRLGLCTSTVQVLTQMPYCLLDHYENRGRLGSODLLNMCQIAKGM 839
Db      774 MAGVSPVYSRLGLCTSTVQVLTQMPYCLLDHYENRGRLGSODLLNMCQIAKGM 833
Qy      840 SYLEDVRLVHRDLAARNVLYKSPNHVKITDGLARLLDIDETEHADGGKVPKMMALRS 939
Db      834 SYLEDVRLVHRDLAARNVLYKSPNHVKITDGLARLLDIDETEHADGGKVPKMMALRS 933
Qy      900 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPICTIDVYM 959
Db      894 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPICTIDVYM 953
Qy      960 IMVKCMWIDSECRPRFRLVSEFSRMARDPQRFVYIQNEGLGSPASPLDSTFYRSLLEDD 1019
Db      954 IMVKCMWIDSECRPRFRLVSEFSRMARDPQRFVYIQNEGLGSPASPLDSTFYRSLLEDD 1013
Qy      1020 MGDLDVABEYLVPQGFPCPDPAAGAGMVRHRRSSSTRSGGDDLLGLPSEEBAPS 1079
Db      1014 MGDLDVABEYLVPQGFPCPDPAAGAGMVRHRRSSSTRSGGDDLLGLPSEEBAPS 1073
Qy      1080 PLABEGAGSDVFPQDGLGMAKAGLSLPHDPSFLQRYSEDETPVLPSTGDIYVAPLTC 1139
Db      1074 PLABEGAGSDVFPQDGLGMAKAGLSLPHDPSFLQRYSEDETPVLPSTGDIYVAPLTC 1133
Qy      1140 SPOEYVNVQPDVRFQPSPREGLPAARPAAGATLERAKTSLPGKGVYKQVAFGAVEN 1199
Db      1134 SPOEYVNVQPDVRFQPSPREGLPAARPAAGATLERAKTSLPGKGVYKQVAFGAVEN 1193
Qy      1200 PEYLTPOGGAAPQHPHPPAPAFDNIYYWDODPERGAPSTFFKGTPTAENEYIGLDV 1259
Db      1194 PEYLTPOGGAAPQHPHPPAPAFDNIYYWDODPERGAPSTFFKGTPTAENEYIGLDV 1253
Qy      1260 PV 1261
Db      1254 PV 1255

RESULT 5
AAW01111
ID      AAW01111 standard; Protein; 1255 AA.
AC      AAW01111;
DT      01-JAN-1997 (first entry)
XX      HER-2/neu protein.
XX      HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW      breast cancer; ovary cancer; colon cancer; lung cancer;
XX      prostate cancer; immunisation; tumour; vaccine; vector.
OS      Homo sapiens.
FH      Key
FT      Domain
FT      Location/Qualifiers
XX      /label= "intracellular domain
XX      /note= "claimed domain, useful for immunisation"
FN      W09630514-A1.
XX      03-OCT-1996.
XX

```

```

PF      28-MAR-1996; 96WO-US01689.
XX      31-MAR-1995; 95US-0414417.
ER      (UNIM ) UNIV WASHINGTON.
XX      Cheever MA, Disis ML;
XX      WPI; 1996-455361/45.
DR      N-PSDB; AAT40739.
XX      DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
PT      treatment of malignancies with which the HER-2/neu oncogene is
PT      associated
XX      Claim 2; Page 56-61; 71pp; English.
PS      Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
XX      the product of the HER-2/neu oncogene (see also AAT40739). The
CC      protein is over-expressed in various cancers, including breast,
CC      ovarian, colon, lung and prostate. The intracellular domain of the
CC      protein can be used to immunise an animal against a malignancy with
CC      which the oncogene is associated. The polypeptide can be produced
CC      in transformed host cells for use in immunisation. Alternatively,
CC      animal cells are transfected in vivo or ex vivo with a viral vector
CC      that directs expression of the polypeptide.
XX      SQ
Sequence 1255 AA:
Query Match 97.1%; Score 6655; DB 17; Length 1255;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1229; Conservative 11; Mismatches 14; Indels 8; Gaps 2
Qy      1 MELAAALRWGILLALALPGAASTQVCTGDMKRLPASPTHIDMLRHLYQGCVQVQNL 60
Db      1 MELAAALRWGILLALALPGAASTQVCTGDMKRLPASPTHIDMLRHLYQGCVQVQNL 60
Qy      61 ELTYLPTNASLSFLODIOEVQGVYLIANQVROVPLQRIYRGVTLFEDNYALAVLDNG 120
Db      61 ELTYLPTNASLSFLODIOEVQGVYLIANQVROVPLQRIYRGVTLFEDNYALAVLDNG 120
Qy      121 DPLNNTPVVAGSPGGIRELQSLSTLEILKGGVLIQRNPQLCYQDTILLKDI FHKXNQLA 180
Db      121 DPLNNTPVVAGSPGGIRELQSLSTLEILKGGVLIQRNPQLCYQDTILLKDI FHKXNQLA 180
Qy      181 LTLIDITNSRACHPCSPMCKGSRCMGSESSDCSLTRTYCAGGACRCKGPLPTDCCHQC 240
Db      181 LTLIDITNSRACHPCSPMCKGSRCMGSESSDCSLTRTYCAGGACRCKGPLPTDCCHQC 240
Qy      241 AAGCTGPKHSDCCLACHFNHSGICEJHCPALVTVNTDTESMPNREGRYTFGASCVTACP 300
Db      241 AAGCTGPKHSDCCLACHFNHSGICEJHCPALVTVNTDTESMPNREGRYTFGASCVTACP 300
Qy      301 YNYLSTDVGSCTLVCPFLHNOEVYAEADGTQRCCKSPCARVCYGLGMFNNFTVSMFLR-V 359
Db      301 YNYLSTDVGSCTLVCPFLHNOEVYAEADGTQRCCKSPCARVCYGLGMFNNFTVSMFLR-V 359
Qy      360 PKVSASHLEFFACCKKIFGSLAFLPSPGDDPASNTAPQYIRANSKFTIGTELTYLVIS 419
Db      360 PKVSASHLEFFACCKKIFGSLAFLPSPGDDPASNTAPQYIRANSKFTIGTELTYLVIS 419
Qy      420 AMPDSLPLDSVPQNLQVIRGRIIHNQAVSLTQIGISWLGRLRELGGSLALHNHTH 479
Db      420 AMPDSLPLDSVPQNLQVIRGRIIHNQAVSLTQIGISWLGRLRELGGSLALHNHTH 479
Qy      474 LCFVHTVPMWQLEFRNHQALLHTARPEDECVGEGLAQCQLCARHCHGPGPTQCVNSQ 533
Db      474 LCFVHTVPMWQLEFRNHQALLHTARPEDECVGEGLAQCQLCARHCHGPGPTQCVNSQ 533
Qy      540 ELAQGVCEBECRYLQGLPREYVNAHRCCLPCHBECQFONGSVYTCFGPEADQCVACHYKDP 599
Db      534 FLRGQCEBECRYLQGLPREYVNAHRCCLPCHBECQFONGSVYTCFGPEADQCVACHYKDP 593

```



```

QY 600 PCVAPCPSPGVKPDLSYPIKPFDEBEGACPCPCINCTHSCVDLIDDKGCAEORASPLTS 659
DB 594 PCVAPCPSPGVKPDLSYPIKPFDEBEGACPCPCINCTHSCVDLIDDKGCAEORASPLTS 653
QY 660 IVSAVVGILLVVLGVVGGILLIKRQOKIRKYTMRELLLOEELVEPLITPSGANPNQAQR 719
DB 654 IISAVGILLVVLGVVGGILLIKRQOKIRKYTMRELLLOEELVEPLITPSGANPNQAQR 713
QY 720 ILKETELRKVKVLSGAGAGTYKGIWIPDGENVKI PVAIKVLRNTPSKANKELIDEAAY 779
DB 714 ILKETELRKVKVLSGAGAGTYKGIWIPDGENVKI PVAIKVLRNTPSKANKELIDEAAY 773
QY 780 MAGVSPYVSRLLGLCLTSTVQVLTQMLPFGCLLDHRENRGRGSGDLLMCKQIAKGM 839
DB 774 MAGVSPYVSRLLGLCLTSTVQVLTQMLPFGCLLDHRENRGRGSGDLLMCKQIAKGM 833
QY 840 SYLEDVRLVHRDLAARNVLYKSPNHVKITDPGLARLLIDETEXHADGKVPKMMALSS 899
DB 834 SYLEDVRLVHRDLAARNVLYKSPNHVKITDPGLARLLIDETEXHADGKVPKMMALSS 893
QY 900 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVTM 959
DB 894 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVTM 953
QY 960 IMVKCMIDSECRPFRELVEFSRMARDPQRFVVIQNEIDLGPASPLDSTFFYSLLLEDD 1019
DB 954 IMVKCMIDSECRPFRELVEFSRMARDPQRFVVIQNEIDLGPASPLDSTFFYSLLLEDD 1013
QY 1020 MGDLYDAEYVLPQGFPCDPAPAGAGMHHRRSSSTSGGGDLTLGLEPSEEBAPRS 1079
DB 1014 MGDLYDAEYVLPQGFPCDPAPAGAGMHHRRSSSTSGGGDLTLGLEPSEEBAPRS 1073
QY 1080 PLAPSEGGSDVPFGDILGMGAKGLSLPHDPSPLORVSEDDTVLPSTTDGYVAPLTC 1139
DB 1074 PLAPSEGGSDVPFGDILGMGAKGLSLPHDPSPLORVSEDDTVLPSTTDGYVAPLTC 1133
QY 1140 SPOEYVNPDPVRPQPPSPREGPLPAARPAGATLERAKT.LSPKNGVYVQVAFGAVEN 1199
DB 1134 SPOEYVNPDPVRPQPPSPREGPLPAARPAGATLERAKT.LSPKNGVYVQVAFGAVEN 1193
QY 1200 PEYLTPQGAAPQHPHPAPSPADNLIYNDOPPEBAGPSPFKGTPTAENDEYGLDV 1259
DB 1194 PEYLTPQGAAPQHPHPAPSPADNLIYNDOPPEBAGPSPFKGTPTAENDEYGLDV 1253
QY 1260 PV 1261
DB 1254 PV 1255

```

```

RESULT 6
AAW92406
ID AAW92406 standard; Protein; 1255 AA.
AC AAW92406;
XX
XX
XX 21-APR-1999 (first entry)
XX
XX Human HER-2/neu oncogene protein.
XX
XX HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
XX malignancy; treatment; tumour.
XX
XX Homo sapiens.
XX
XX Key
XX Region 676..1255
XX /note="region which elicits immune response"
XX
XX US5869445-A.
XX
XX 09-FEB-1999.
XX
XX 01-APR-1996; 96US-0625101.

```

```

XX 01-APR-1996; 96US-0625101.
PR 17-MAR-1993; 93US-0033644.
PR 12-AUG-1993; 93US-0106112.
PR 31-MAR-1995; 95US-0414417.
XX
XX (UNIM ) UNIV WASHINGTON.
XX
XX Cheever MA, Disis ML;
XX
XX MPI: 1999-152835/13.
XX N-PsDB; AAX01912.
XX
XX Use of HER-2/neu polypeptides - for eliciting an immune response to
XX an HER-2/neu associated malignancy, particularly for treating or
XX preventing tumours
XX
XX Claim 3; Column 31-38; 26pp; English.
XX
XX This sequence represents the human HER-2/neu oncogene protein. A fragment
XX of this protein is used in a method for eliciting or enhancing an immune
XX response to HER-2/neu protein. The polypeptide can stimulate T cells and
XX B cells to produce an immune response to the HER-2/neu protein. The
XX method can be used for immunisation against a malignancy in which the
XX HER-2/neu oncogene is associated and in the treatment of an existing
XX tumour, or to prevent tumour occurrence or recurrence.
XX
XX Sequence 1255 AA;
XX
XX Query Match 97.1%; Score 6655; DB 20; Length 1255.
XX Best Local Similarity 97.4%; Pred. No. 0;
XX Matches 1229; Conservative 11; Mismatches 14; Indels 8; Gaps 2

```

```

QY 1 MELALCRWGLLALLPPGAASVQVGTGDMKRLPASPETHLDMRLHYQGGVQGNL 60
DB 1 MELALCRWGLLALLPPGAASVQVGTGDMKRLPASPETHLDMRLHYQGGVQGNL 60
QY 61 ELYLPNNAISLFIQDIQEVQVYLLAHNQRQVPLQRLRYRGTLFEDNVALAVLNG 120
DB 61 ELYLPNNAISLFIQDIQEVQVYLLAHNQRQVPLQRLRYRGTLFEDNVALAVLNG 120
QY 121 DPLNNTPTVPGASPGGLREIQLSLTEILKGVLLIQRNPQLCYQDTILMKDIFHKNNQLA 180
DB 121 DPLNNTPTVPGASPGGLREIQLSLTEILKGVLLIQRNPQLCYQDTILMKDIFHKNNQLA 180
QY 181 LTLIDINRSRACHPCSPCKGRKCGWESSSEDCSLRTVCAAGCARCKGPLPTDCCHQC 240
DB 181 LTLIDINRSRACHPCSPCKGRKCGWESSSEDCSLRTVCAAGCARCKGPLPTDCCHQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYVNTDTESMNPREGRTTFGASCYTTAC 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYVNTDTESMNPREGRTTFGASCYTTAC 300
QY 301 YNYLSTDVGSCTLVCPLNQEVLAEDGTQRCCKSPCARVCGYGLMFNFTVSEFLR-V 359
DB 301 YNYLSTDVGSCTLVCPLNQEVLAEDGTQRCCKSPCARVCGYGLMFNFTVSEFLR-V 359
QY 360 PKYSASHLERFACCKKIFGSLATLPSPFGDDPASNTAPQYIRANSKFQITLTGLYIS 419
DB 360 PKYSASHLERFACCKKIFGSLATLPSPFGDDPASNTAPQYIRANSKFQITLTGLYIS 419
QY 420 AMPDSLPLDSVFQNTLVIRGRILHNAAYSLTQGLGISWLGRLRELGSGALHHNTH 479
DB 420 AMPDSLPLDSVFQNTLVIRGRILHNAAYSLTQGLGISWLGRLRELGSGALHHNTH 479
QY 480 LCFVHTVPMQDLFRNPHQALLHTANPEDECVGEGIALCHQLCARGHCHWPGPTQCVNSQ 539
DB 480 LCFVHTVPMQDLFRNPHQALLHTANPEDECVGEGIALCHQLCARGHCHWPGPTQCVNSQ 539
QY 540 FLRGQECVEECRVLQGLPREYVNAHRCILPCHPEQCPONGSVTCTFGPEADQCVACHAYDP 599
DB 540 FLRGQECVEECRVLQGLPREYVNAHRCILPCHPEQCPONGSVTCTFGPEADQCVACHAYDP 599

```

```
QY 600 PFCVACPSGVKPDLSYMPIMKFPDEEGACOPCPINCTHSCVDLDDKCGPAEQRASPLTS 659
DB 594 PFCVACPSGVKPDLSYMPIMKFPDEEGACOPCPINCTHSCVDLDDKCGPAEQRASPLTS 653
QY 660 IWSAVVGIILLVVLGVVFGIILIKRQOKIKRYTMRLLQETELVEPLTPSGAMNQOMR 719
DB 654 IISAVVGIILLVVLGVVFGIILIKRQOKIKRYTMRLLQETELVEPLTPSGAMNQOMR 713
QY 720 ILKETELRKVKTLGSAFGTYVKGIMIPGENVKI PVAKYLRENTSGKANKELIDEAYV 779
DB 714 ILKETELRKVKTLGSAFGTYVKGIMIPGENVKI PVAKYLRENTSGKANKELIDEAYV 773
QY 780 MAGVSPYVSRLLGLICLTSTVQLVTLQMPYGCILDHVENNGRLGSDQLLWMCQIAKGM 839
DB 774 MAGVSPYVSRLLGLICLTSTVQLVTLQMPYGCILDHVENNGRLGSDQLLWMCQIAKGM 833
QY 840 SYLEDVRLVHRDLAANVLYKSPNNVKTIDFSLAALLIDETETHADGKVPIMWMALES 899
DB 834 SYLEDVRLVHRDLAANVLYKSPNNVKTIDFSLAALLIDETETHADGKVPIMWMALES 893
QY 900 ILRRRFTHSDVMSYGVTVMEIMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYM 959
DB 894 ILRRRFTHSDVMSYGVTVMEIMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYM 953
QY 960 IYVKCMMIDSECRPRPRELVSFSFMAADPPQFVYIQNEDIJGPA SPLDSTFYRSLLDDDD 1019
DB 954 IYVKCMMIDSECRPRPRELVSFSFMAADPPQFVYIQNEDIJGPA SPLDSTFYRSLLDDDD 1013
QY 1020 MGDLYDAEEYLVPOOGFCPCDPAPAGAGMHHRRSSSTSGGGDLTLGLEPSEEEAPRS 1079
DB 1014 MGDLYDAEEYLVPOOGFCPCDPAPAGAGMHHRRSSSTSGGGDLTLGLEPSEEEAPRS 1073
QY 1080 PLAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPDJORYSDDPTVLPJSEDTGYVAPLTC 1139
DB 1074 PLAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPDJORYSDDPTVLPJSEDTGYVAPLTC 1133
QY 1140 SPOPEYVNOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGYKCVFAFGAVEN 1199
DB 1134 SPOPEYVNOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGYKCVFAFGAVEN 1193
QY 1200 PEYLPPOGGAAPQHPHPAFAFPANLYYWDODPPERGAPSTFKGTPTAENPEYLGLDV 1259
DB 1194 PEYLPPOGGAAPQHPHPAFAFPANLYYWDODPPERGAPSTFKGTPTAENPEYLGLDV 1253
QY 1260 PV 1261
DB 1254 PV 1255

RESULT 7
AAB21198
ID AAB21198 standard; protein; 1255 AA.
XX
AC AAB21198;
XX
DT 12-JAN-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
KW Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW colon cancer.
XX
OS Homo sapiens.
XX
PN MO200044899-A1.
PD 03-AUG-2000.
XX
PF 28-JAN-2000; 2000MO-US02164.
XX
PR 29-JAN-1999; 99US-0117976.
XX
```

```
PA (COR-) CORIXA CORP.
PA (SMK) SMITHKLINE BEECHAM.
XX
PI Cheever MA, Gheysen D;
XX
DR WPI: 2000-505976/45.
DR N-PSDB; AAA89736.
XX
PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PS prostate cancers -
XX
PS Claim 52; Fig 7; 128pp; English.
XX
CC The present sequence is the human HER-2/neu protein. It is a member of
CC the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against
CC these neoplasias.
XX
SQ Sequence 1255 AA;
QY Query Match 97.1%; Score 665; DB 21; Length 1255;
QY Best Local Similarity 97.4%; Pred. No. 0;
QY Matches 1229; Conservative 11; Mismatches 14; Indels 8; Gaps ;
QY 1 METALCMWGLLALPPGAASTOVCTGTMKRLPASPETHLDMRLHYGCGVVGNTL 60
QY 1 METALCMWGLLALPPGAASTOVCTGTMKRLPASPETHLDMRLHYGCGVVGNTL 60
DB
QY 61 ELTYLPTNASLSFLDIDGEOGVYLAHNOYRQVPLQRLRVGTQFEDNYALAVLNDG 126
DB 61 ELTYLPTNASLSFLDIDGEOGVYLAHNOYRQVPLQRLRVGTQFEDNYALAVLNDG 126
QY 121 DPLNNTTPTVGTASRPGAELELDELTELKGVYIQSNPQCYODTILMWDIFKKNQLA 186
DB 121 DPLNNTTPTVGTASRPGAELELDELTELKGVYIQSNPQCYODTILMWDIFKKNQLA 186
QY 181 LTLIDTNSRACHPSCPMCKSRCKGESSDQSLRTVCAGCARGCPPTDCHEQC 246
DB 181 LTLIDTNSRACHPSCPMCKSRCKGESSDQSLRTVCAGCARGCPPTDCHEQC 246
QY 241 AAGCTGPRGSDCLAHNHSIGIELCPALVTYNTDTPESMPNBERYTFGASCVTACP 306
DB 241 AAGCTGPRGSDCLAHNHSIGIELCPALVTYNTDTPESMPNBERYTFGASCVTACP 306
QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGEMNFTVSFWLR-V 355
DB 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGEMNFTVSFWLR-V 355
QY 360 PKVASHLLEEFAGCKKIFGSLAFLPESFDGDPASNTAPQYIKANSKFIGTEETGYLYIS 415
DB 360 PKVASHLLEEFAGCKKIFGSLAFLPESFDGDPASNTAPQYIKANSKFIGTEETGYLYIS 415
QY 420 AMPDCLPPLSVFQNLQVIRGRILNKGYSYLTLOGLSIMGLSLRSLRGLLIIHNTH 479
DB 420 AMPDCLPPLSVFQNLQVIRGRILNKGYSYLTLOGLSIMGLSLRSLRGLLIIHNTH 479
QY 444 AMPDCLPPLSVFQNLQVIRGRILNKGYSYLTLOGLSIMGLSLRSLRGLLIIHNTH 473
DB 444 AMPDCLPPLSVFQNLQVIRGRILNKGYSYLTLOGLSIMGLSLRSLRGLLIIHNTH 473
QY 480 LCFVHTVPMQLFNNPQALHTANREDECVGEGALCHQLCARGHGCGPPTQCVNCSQ 539
DB 474 LCFVHTVPMQLFNNPQALHTANREDECVGEGALCHQLCARGHGCGPPTQCVNCSQ 533
QY 540 FLRGQECVEECRVYQGLPREYVNAHCLPCHPEQPNQNGSVTTFGEBADQVCAHYKXP 599
DB 534 FLRGQECVEECRVYQGLPREYVNAHCLPCHPEQPNQNGSVTTFGEBADQVCAHYKXP 593
QY 600 PFCVACPSGVKPDLSYMPIMKFPDEEGACOPCPINCTHSCVDLDDKCGPAEQRASPLTS 659
```

```

Db      594 PFCVAPCSGVKPDLSYMPIMKFPDEEGACQPCPINCTHSCVDLDKGPAGQASPLTS 653
QY      660 IVSAVVGILVVYGVGVGILIKRQOKIRKYIMRRLQSTELVEPLTPSGAMPQNAQNR 719
Db      654 IISAVGILLVVYGVGVGILIKRQOKIRKYIMRRLQSTELVEPLTPSGAMPQNAQNR 713
QY      720 ILKETELRKVKVLGSGAGFVYKGIWIPDGENVKIPAIIVLRNTPSKANKELDEAVY 779
Db      714 ILKETELRKVKVLGSGAGFVYKGIWIPDGENVKIPAIIVLRNTPSKANKELDEAVY 773
QY      780 MAGVSPVYSRLILGICLTSTVQVLTQMPYGLLDHRENRGLSGODLLNMCQIAKGM 839
Db      774 MAGVSPVYSRLILGICLTSTVQVLTQMPYGLLDHRENRGLSGODLLNMCQIAKGM 833
QY      840 SYLEDVRLVHBDLAARNVLYKSPNHVKITDFGLARLLIDETEVHADGGKVPKXMALES 899
Db      834 SYLEDVRLVHBDLAARNVLYKSPNHVKITDFGLARLLIDETEVHADGGKVPKXMALES 893
QY      900 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGRLLPQPPICITDVTM 959
Db      894 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGRLLPQPPICITDVTM 953
QY      960 IMVCKMIDSECRFRFELVSEFSRMARDPQRFVLIQNEIDLGPASPLDSTFYRSLLDSD 1019
Db      954 IMVCKMIDSECRFRFELVSEFSRMARDPQRFVLIQNEIDLGPASPLDSTFYRSLLDSD 1013
QY      1020 MGDIVDAEYLVPOQGFCCPDPAAGAGMVAHHRSSSTRSGGDLTLGLEPSEEBAPRS 1079
Db      1014 MGDIVDAEYLVPOQGFCCPDPAAGAGMVAHHRSSSTRSGGDLTLGLEPSEEBAPRS 1073
QY      1080 PLASEGAGSVYFPGDILGMAAKGLQSLPHDPSPLQRYSEDVTVPPESTQDYAAPLTC 1139
Db      1074 PLASEGAGSVYFPGDILGMAAKGLQSLPHDPSPLQRYSEDVTVPPESTQDYAAPLTC 1133
QY      1140 SPQPEYVNPQDVRQPPSPREGPLPAARPAATLERAKTLSPKNGVYVDVAFGGAIVEN 1199
Db      1134 SPQPEYVNPQDVRQPPSPREGPLPAARPAATLERAKTLSPKNGVYVDVAFGGAIVEN 1193
QY      1200 PEYLTPOGGAAPQHPHPAPSPAFDNLVYWDQDPPERGAPESTFKGTPTAENPEYIGLDV 1259
Db      1194 PEYLTPOGGAAPQHPHPAPSPAFDNLVYWDQDPPERGAPESTFKGTPTAENPEYIGLDV 1253
QY      1260 PV 1261
Db      1254 PV 1255

RESULT 8
AAy84780
ID      AAy84780 standard; Protein; 1255 AA.
XX
AC      AAy84780;
XX
DT      08-AUG-2000 (first entry)
XX
DE      Amino acid sequence of the SPLICE erbB-2 receptor protein.
XX
KW      SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
KW      tumor cell proliferation; tissue degeneration; arthropathy;
KW      bone resorption; inflammatory disease; degenerative disorder;
KW      wound healing.
XX
OS      Homo sapiens.
XX
PN      W0200020579-A1.
XX
PD      13-APR-2000.
XX
PF      01-OCT-1999; 99MO-CA00912.
XX
PR      02-OCT-1998; 98US-01.65192.
XX

```

```

PA      (UYWC-) UNIV MCMMASTER.
XX
PI      Muller WJ, Siegel PM;
XX
DR      WPI: 2000-303768/26.
XX
DR      N-PSDB; AAA14812.
XX
PT      Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
XX
PS      erbB-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS      Claim 3; Fig 2; 60pp; English.
XX
CC      The present sequence represents a SPLICE erbB-2 receptor protein. The
CC      protein has an in-frame deletion of 16 amino acids, 2 of which are
CC      conserved cysteine residues, compared to the unspliced protein. The
CC      erbB-2 polynucleotide is used to construct probes for detecting
CC      disorders of cell transformation such as cancer. Antibodies to the
CC      protein may be used to detect SPLICE erbB-2 in a sample. Agents
CC      (e.g. antisense oligonucleotides) which inhibit the expression of
CC      SPLICE erbB-2 are useful for reducing tumor cell proliferation and
CC      treating cancer. Substances which stimulate SPLICE erbB-2 are useful
CC      for treating conditions involving damaged cells including conditions
CC      in which degeneration of tissue occurs, such as arthropathy, bone
CC      resorption, inflammatory diseases, degenerative disorders of the
CC      central nervous system and wound healing.
XX
SQ      Sequence 1255 AA;
XX
Query Match 97.1%; Score 6655; DB 21; Length 1255;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1229; Conservative 11; Mismatches 14; Indels 8; Gaps 2
QY      1 MELALGRWGILLALPPGAASTQVCTGDMKRLPASPEHLMRLHYOGQGVQGNL 60
Db      1 MELALGRWGILLALPPGAASQVCTGDMKRLPASPEHLMRLHYOGQGVQGNL 60
QY      61 ELTYLPNNAISLFIQDIQEVQGVYVLAHNOVROPVQRLRIYRGTLFEDNVALAVLNG 120
Db      61 ELTYLPNNAISLFIQDIQEVQGVYVLAHNOVROPVQRLRIYRGTLFEDNVALAVLNG 120
QY      121 DPLNNTPTVYGASPGGRLRELQSLLEILKAGVLIQRNPOLCYQDTILMKDIFHKNNQLA 180
Db      121 DPLNNTPTVYGASPGGRLRELQSLLEILKAGVLIQRNPOLCYQDTILMKDIFHKNNQLA 180
QY      181 LTLIDINRSAPCHPCSEMKGSRGWESSEDCQSLRTVYQAGS ZARCKGPLPTCCHEQC 240
Db      181 LTLIDINRSAPCHPCSEMKGSRGWESSEDCQSLRTVYQAGS ZARCKGPLPTCCHEQC 240
QY      241 AAGCTGPHSDCLALCHFNHSGICEIHCPALVYVNDTFESNPNBGRYTFGASCTYACP 300
Db      241 AAGCTGPHSDCLALCHFNHSGICEIHCPALVYVNDTFESNPNBGRYTFGASCTYACP 300
QY      301 YNYLSTDVGSCTVLCPLHNOEVAEDGTQRCCKSPKCARVVCYGLMFNNFTVSWLR-V 359
Db      301 YNYLSTDVGSCTVLCPLHNOEVAEDGTQRCCKSPKCARVVCYGLMFNNFTVSWLR-V 359
QY      360 PKYSASLLEFPACCKIFGSLAFLPSPFGDPAASNTAPQYIKANSKFTGITEULGYLYIS 419
Db      360 PKYSASLLEFPACCKIFGSLAFLPSPFGDPAASNTAPQYIKANSKFTGITEULGYLYIS 419
QY      420 AMPDSLPLDLVFNQLOVYIRGRILAHNAYSLTQGLGISWLGSRSLREIGSGALAHNNTH 479
Db      420 AMPDSLPLDLVFNQLOVYIRGRILAHNAYSLTQGLGISWLGSRSLREIGSGALAHNNTH 479
QY      414 AMPDSLPLDLVFNQLOVYIRGRILAHNAYSLTQGLGISWLGSRSLREIGSGALAHNNTH 473
Db      414 AMPDSLPLDLVFNQLOVYIRGRILAHNAYSLTQGLGISWLGSRSLREIGSGALAHNNTH 473
QY      480 LCPVHTVPMQDLFRNPHQALLHTANRPEDECVGEGSLAQHLCARHGCMWPGPTQCVNCSQ 539
Db      474 LCPVHTVPMQDLFRNPHQALLHTANRPEDECVGEGSLAQHLCARHGCMWPGPTQCVNCSQ 533
QY      540 FLRGQCEVEECRYLQGLPREYVNAARCLCHBEQCPQNSSTYTCFPEALQCYACAHYNDP 599
Db      534 FLRGQCEVEECRYLQGLPREYVNAARCLCHBEQCPQNSSTYTCFPEALQCYACAHYNDP 593
QY      600 PFCVAPCSGVKPDLSYMPIMKFPDEEGACQPCPINCTHSCVDLDKGPAGQASPLTS 659

```

```
Db 594 PFCVACPSGVKPDLSYMPIMKFPDEBACQPCPCINCHSCVDLDDGCCAEGASPLTS 653
Qy 660 IVSAVVGILLVVLGVGVGILLIKRQCKIRKRYTRRLLOETELVEPLTPSGAMPNQAQMR 719
Db 654 IISAVVGLLVVLGVGVGILLIKRQCKIRKRYTRRLLOETELVEPLTPSGAMPNQAQMR 713
Qy 720 IKETELRKVKVLGSGAGFTYKGIWIDGENVMI PVAIKYLENTSPKANKELDEAYV 779
Db 714 IKETELRKVKVLGSGAGFTYKGIWIDGENVMI PVAIKYLENTSPKANKELDEAYV 773
Qy 780 MAGVSPYVSRLLGLITSTVQLVLTQMPYQCLLDHYRENRGRIGSQDLLNMCQIAKGM 839
Db 774 MAGVSPYVSRLLGLITSTVQLVLTQMPYQCLLDHYRENRGRIGSQDLLNMCQIAKGM 833
Qy 840 SYLEDVRLVHRDLAARNLVKSPNHVKITDGLARLLDIDETEHADGGVPIKMALES 899
Db 834 SYLEDVRLVHRDLAARNLVKSPNHVKITDGLARLLDIDETEHADGGVPIKMALES 893
Qy 900 ILRRFTHQSDVWSYGVTVWELMTFGANPYDGI PAREIPDLEKGERLPQRPICITIDVYM 959
Db 894 ILRRFTHQSDVWSYGVTVWELMTFGANPYDGI PAREIPDLEKGERLPQRPICITIDVYM 953
Qy 960 IMVKCMITDSECRPRFRELVESEFARMADPORFVVIQNEDLGPRASPLDSTFYSLLEDD 1019
Db 954 IMVKCMITDSECRPRFRELVESEFARMADPORFVVIQNEDLGPRASPLDSTFYSLLEDD 1013
Qy 1020 MGDLDVAEEYLVPQGGFCPPDPAPAGAGMWHRRSSSTRSGGDLTLGLEPSEEEAPRS 1079
Db 1014 MGDLDVAEEYLVPQGGFCPPDPAPAGAGMWHRRSSSTRSGGDLTLGLEPSEEEAPRS 1073
Qy 1080 PLABSEGASDVFDGDLGMGAAGKIQSLPTHDPSPLORYSEDPVPLPSTDTGYVALTIC 1139
Db 1074 PLABSEGASDVFDGDLGMGAAGKIQSLPTHDPSPLORYSEDPVPLPSTDTGYVALTIC 1133
Qy 1140 SPQPEYVNPQPVRRPOPSPREGRPLPAARPAAGATLERAKTILSPGKNGVVKVPAFGAVEN 1199
Db 1134 SPQPEYVNPQPVRRPOPSPREGRPLPAARPAAGATLERAKTILSPGKNGVVKVPAFGAVEN 1193
Qy 1200 PEYLTPOCGAAPQHPPPAFSPAFDNLTYWDQDPERGAPPSTFKGPTAENPEYGLDV 1259
Db 1194 PEYLTPOCGAAPQHPPPAFSPAFDNLTYWDQDPERGAPPSTFKGPTAENPEYGLDV 1253
Qy 1260 PV 1261
Db 1254 PV 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
XX
AC AAB85458;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
KW Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
XX oncogene; cancer; cytosolic; vaccine; p185; C-erbB2.
XX
OS Homo sapiens.
XX
PN WO200153463-A2.
XX
PD 26-JUL-2001.
XX
PF 19-JAN-2001; 2001WO-US01850.
XX
PR 21-JAN-2000; 2000US-0177545.
XX
PA (CORI-) CORIXA CORP.
XX
```

```
PI Cheever MA, Hand-Zimmermann S;
XX WPI: 2001-476112/51.
DR N-PSDB; AAB23392.
XX
PT New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
PT for treating or preventing cancer, e.g. breast cancer
XX
PS Claim 2; Page 41-46; 49pp; English.
XX
CC The invention provides an isolated antigen-presenting cell, which
CC expresses at least an immunogenic portion of a polypeptide that produces
CC an immune response to HER-2/neu protein. The antigen-presenting cells are
CC useful as vaccines for eliciting or enhancing an immune response to
CC HER-2/neu protein, particularly in treating or preventing malignancies in
CC which the HER-2/neu oncogene is associated. Specifically, these are
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC colon, lung or prostate cancers. The present sequence represents
CC the human HER-2/neu protein (also known as p185 or C-erbB2).
XX
SQ Sequence 1255 AA;
Qy Query Match 97.1%; Score 6655; DB 22; Length 1255;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1229; Conservative 11; Mismatches 14; Indels 8; Gaps 2
Qy 1 MELALACWGLLALLLPAGASTVCTGTAKRLPASPETHIDMLRHLYQGQGVQGNL 60
Db 1 MELALACWGLLALLLPAGASTVCTGTAKRLPASPETHIDMLRHLYQGQGVQGNL 60
Qy 61 ELTYLPTNASLSPLODIOEVQGVYLAAHNOVQVPLQRLRIVRGTOFEDNYVALVDNG 120
Db 61 ELTYLPTNASLSPLODIOEVQGVYLAAHNOVQVPLQRLRIVRGTOFEDNYVALVDNG 120
Qy 121 DLNNTTPYVAGASGGARELQSLSTLILKGVILQGNPQLCYODPTILWMDIFHKNQOLA 180
Db 121 DLNNTTPYVAGASGGARELQSLSTLILKGVILQGNPQLCYODPTILWMDIFHKNQOLA 180
Qy 121 DPLNNTTPYVAGASGGARELQSLSTLILKGVILQGNPQLCYODPTILWMDIFHKNQOLA 180
Db 121 DPLNNTTPYVAGASGGARELQSLSTLILKGVILQGNPQLCYODPTILWMDIFHKNQOLA 180
Qy 181 LTLIDTNRSRACHPCSMCKSGRWGSSSDCCSLRTVCAAGACACKGLPTDCHEQC 240
Db 181 LTLIDTNRSRACHPCSMCKSGRWGSSSDCCSLRTVCAAGACACKGLPTDCHEQC 240
Qy 241 AAGCTGPRHSDCLACHFNHSGICEHLCPALVYNTDTPFSMNPGRGYTFGASCYTACP 300
Db 241 AAGCTGPRHSDCLACHFNHSGICEHLCPALVYNTDTPFSMNPGRGYTFGASCYTACP 300
Qy 301 YNYLSTDVGSCTLVCPHNOEVTAEDEGTORCEKSCPKARVCYGLMFPNFTVSPWLR-V 359
Db 301 YNYLSTDVGSCTLVCPHNOEVTAEDEGTORCEKSCPKARVCYGLMFPNFTVSPWLR-V 359
Qy 360 PKVASHLPEEPAGCKITFGSLAFIPESFDGDPASNTAPQYIKANSKFIGITELGYLYIS 419
Db 360 PKVASHLPEEPAGCKITFGSLAFIPESFDGDPASNTAPQYIKANSKFIGITELGYLYIS 413
Qy 420 AMPDSLPLASVFOQLVYIRGRIILHNGVYSLTLGGLGISMWGLASLRELISGLALIHNNH 479
Db 420 AMPDSLPLASVFOQLVYIRGRIILHNGVYSLTLGGLGISMWGLASLRELISGLALIHNNH 473
Qy 474 LCPVHTVPMQDLFENPHQALLHTANREDECEVGEGLACHQLCARGHGCMWGGPPTCVNCSQ 533
Db 474 LCPVHTVPMQDLFENPHQALLHTANREDECEVGEGLACHQLCARGHGCMWGGPPTCVNCSQ 533
Qy 540 FLRGQECVEECRYLQGLPREYVNAARHCLPHREPCQONSGSYTFGEBAOQCVAACAYKXP 599
Db 540 FLRGQECVEECRYLQGLPREYVNAARHCLPHREPCQONSGSYTFGEBAOQCVAACAYKXP 593
Qy 600 PFCVACPSGVKPDLSYMPIMKFPDEBACQPCPCINCHSCVDLDDGCCAEGASPLTS 659
Db 594 PFCVACPSGVKPDLSYMPIMKFPDEBACQPCPCINCHSCVDLDDGCCAEGASPLTS 653
Qy 660 IVSAVVGILLVVLGVGVGILLIKRQCKIRKRYTRRLLOETELVEPLTPSGAMPNQAQMR 719
Db 654 IISAVVGLLVVLGVGVGILLIKRQCKIRKRYTRRLLOETELVEPLTPSGAMPNQAQMR 713
```

```

Db      654  IISAVVGLLVVVLGVFGILLIKRRQOKIRKXVTKMRLLQETELVEPLTPSGAMENCAQMR 713
Qy      720  IKETELRKVKVLSGAGFTVYKGIIMPDGENVKIPVAIKYLRNTSPKANKETIDEAYV 779
Db      714  IKETELRKVKVLSGAGFTVYKGIIMPDGENVKIPVAIKYLRNTSPKANKETIDEAYV 773
Qy      780  MAGVSPVYSRLIGICLTSTVOLVTLQMPYGGCLDHVRENGRLGSGODLLMCMQIAKM 839
Db      774  MAGVSPVYSRLIGICLTSTVOLVTLQMPYGGCLDHVRENGRLGSGODLLMCMQIAKM 833
Qy      840  SYLEDVRLVHRDLAARNLVKSPNHVKITDREGLAETLIDETEVHADGCKPIKMALES 899
Db      834  SYLEDVRLVHRDLAARNLVKSPNHVKITDREGLAETLIDETEVHADGCKPIKMALES 893
Qy      900  ILRRRFTHQSDVMSYGVTVWELMTGAKPYDGIAPAREIPDLLEKGERLPQPICTIDVM 959
Db      894  ILRRRFTHQSDVMSYGVTVWELMTGAKPYDGIAPAREIPDLLEKGERLPQPICTIDVM 953
Qy      960  IMVKCMNIDSECRPRRELVSFSSMARDPQRFVYIQNEDLGASPLDSTFYRSLDDDD 1019
Db      954  IMVKCMNIDSECRPRRELVSFSSMARDPQRFVYIQNEDLGASPLDSTFYRSLDDDD 1013
Qy      1020  MGDLDVAEEYLVPOQGFPCPDPAAGAGMHHRRSSSTRSGGDLTIGLEPSEEAPRS 1079
Db      1014  MGDLDVAEEYLVPOQGFPCPDPAAGAGMHHRRSSSTRSGGDLTIGLEPSEEAPRS 1073
Qy      1080  PLAPSEGAGSDVFDGDLGMAAKGILQSLPTDPSFLQYSDDPTVPLPSETDGYVADLTG 1139
Db      1074  PLAPSEGAGSDVFDGDLGMAAKGILQSLPTDPSFLQYSDDPTVPLPSETDGYVADLTG 1133
Qy      1140  SPQPEYVNOVDYRPPSPRREGPLPAAPAGATLERPKTLSPGNGVYKDVFAFGAVEN 1199
Db      1134  SPQPEYVNOVDYRPPSPRREGPLPAAPAGATLERPKTLSPGNGVYKDVFAFGAVEN 1193
Qy      1200  PEYLTPOGGAAPQHPHPAFSPAFDNLVYWDODPERGAPSTFKGPTTAENPEYGLDV 1259
Db      1194  PEYLTPOGGAAPQHPHPAFSPAFDNLVYWDODPERGAPSTFKGPTTAENPEYGLDV 1253
Qy      1260  PY 1261
Db      1254  PY 1255

RESULT 10
AAG88267
ID      AAG88267 standard; Protein; 1255 AA.
AC      AAG88267;
XX
DT      11-SEP-2001 (first entry)
XX
DE      HER2/neu amino acid sequence.
XX
KM      Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
XX      immune response; vaccine; tumour; cancer; cytotoxic; immunostimulant;
XX      tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
OS      Homo sapiens.
XX
PN      MO200141787-A1.
PD      14-JUN-2001.
PF      11-DEC-2000; 2000MO-US33591.
PR      10-DEC-1999; 99US-0458299.
PA      (EPI-M-) EPIMMUNE INC.
PI      Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
DR      Keogh E;
      WPI; 2001-374995/39.
```

XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer -
XX
XX
XX Disclosure; Page 15; 1999p; English.

XX The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I) bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytotoxic
CC and immunostimulant activities, and can be used in vaccines. (I), (II)
CC and (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample from a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention.
XX

SQ Sequence 1255 AA;

Query Match 97.1%; Score 6655; DB 22; Length 1255;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1229; Conservative 11; Mismatches 14; Indels 8; Gaps :

```

Qy      1 MELALCRWGLLLALPPGAASSTVCTGTDMKRLPASPETHLDMRLHYOGCGVQVGNL 60
Db      1 MELALCRWGLLLALPPGAASSTVCTGTDMKRLPASPETHLDMRLHYOGCGVQVGNL 60
Qy      61 ELTYPTNASSLSFLDIOEVQGYVLIHNOVQVPLQRLIRVGTQVLFEDNYALAVLDNG 124
Db      61 ELTYPTNASSLSFLDIOEVQGYVLIHNOVQVPLQRLIRVGTQVLFEDNYALAVLDNG 124
Qy      121 DPLNNTTPTVGTASPGGLRELQRLSTELIKGGVLIQNNPOLCTQDTLLMKDIFRKNOLA 180
Db      121 DPLNNTTPTVGTASPGGLRELQRLSTELIKGGVLIQNNPOLCTQDTLLMKDIFRKNOLA 180
Qy      181 LTLIDTNSRACHPCSPCKSGSRGSSSEDCQSLRTVCAAGGACRCKGPLPDDCCEQC 240
Db      181 LTLIDTNSRACHPCSPCKSGSRGSSSEDCQSLRTVCAAGGACRCKGPLPDDCCEQC 240
Qy      241 AAGCTGPRHSQCLACLFHNSGICELCPALVYNTDFFSMPBERRYFGASCTYACP 300
Db      241 AAGCTGPRHSQCLACLFHNSGICELCPALVYNTDFFSMPBERRYFGASCTYACP 300
Qy      301 YNYLSTDVGSCTLYCPLHNOEVTAEADTQRCCKSPCARVCYGLGWFNFTVSFWLR-V 359
Db      301 YNYLSTDVGSCTLYCPLHNOEVTAEADTQRCCKSPCARVCYGLGWFNFTVSFWLR-V 359
Qy      360 PKVASHLEEFAGCKKIFGSLAPFESFDGDPASNTAPQYIKANSKFIGTETLGYLYIS 419
Db      360 PKVASHLEEFAGCKKIFGSLAPFESFDGDPASNTAPQYIKANSKFIGTETLGYLYIS 419
Qy      420 AMPDLPDLASVFOQLVYIRGRILHNGVYSLTQLGQISWGLSLRELISGALIHNNH 479
Db      414 AMPDLPDLASVFOQLVYIRGRILHNGVYSLTQLGQISWGLSLRELISGALIHNNH 473
Qy      480 LCFVHTVPMQDLFNNHQALLHTANRDECEVGEALCHQQLCARGHGCMWPGPTQVNCGQ 539
Db      480 LCFVHTVPMQDLFNNHQALLHTANRDECEVGEALCHQQLCARGHGCMWPGPTQVNCGQ 539
```

```

Db      474 LCFVHTVPMDOLEFRNPQALLHTANRPEDECVSEGLACHQICARGHCWGPPTQCVCNSQ 533
QY      540 FLRGOECVEECRVLQGLPREYVNAHCLPCHPECCPONGSVTTCGPEADQCVACAHTKDP 539
Db      534 FLRGOECVEECRVLQGLPREYVNAHCLPCHPECCPONGSVTTCGPEADQCVACAHTKDP 593
QY      600 PFCVAPCEGKVPDLSTYMPIMKFPDEGACQCPINCHSCVDLDDKCPAEPQASPLTS 659
Db      594 PFCVAPCEGKVPDLSTYMPIMKFPDEGACQCPINCHSCVDLDDKCPAEPQASPLTS 653
QY      660 IYSAVVGILLVVLGVNIGLILKRRQOKIRKXTMRLLQETELVEPLTPSGAMPNQOMR 719
Db      654 IYSAVVGILLVVLGVNIGLILKRRQOKIRKXTMRLLQETELVEPLTPSGAMPNQOMR 713
QY      720 IKETELRKVKYLSGAGFTVYKGIWIDGENVKIPVAIKYLRNTSPKAKELIDZAYV 779
Db      714 IKETELRKVKYLSGAGFTVYKGIWIDGENVKIPVAIKYLRNTSPKAKELIDZAYV 773
QY      780 MAGVSPVYSRLIGICTSTVOLVTQMLPYGCLDHYENRGRLGSDOLLNMCQIAKGM 839
Db      774 MAGVSPVYSRLIGICTSTVOLVTQMLPYGCLDHYENRGRLGSDOLLNMCQIAKGM 833
QY      840 SYLEDVRLVHRLAARNVLKSPNHVKITDFGLARLIDIDETEVHADGKVPIMKMALES 899
Db      834 SYLEDVRLVHRLAARNVLKSPNHVKITDFGLARLIDIDETEVHADGKVPIMKMALES 893
QY      900 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPPICTIDVYM 959
Db      894 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPPICTIDVYM 953
QY      960 IMVKCMIDSECRPFRELVSFSSMAADPQRFVIONEDIGPASPLDSTYRSLLEDD 1019
Db      954 IMVKCMIDSECRPFRELVSFSSMAADPQRFVIONEDIGPASPLDSTYRSLLEDD 1013
QY      1020 MGDLYDAEYLVPOQGFPCPDPAFGAGMVAHRRSSSTRSGGDLTLGLEPSEEAPRS 1079
Db      1014 MGDLYDAEYLVPOQGFPCPDPAFGAGMVAHRRSSSTRSGGDLTLGLEPSEEAPRS 1073
QY      1080 PLAPSEGAGSDVFDDDLQMGAAKIGISPTHTDPSFLQYSEDFTVPLPSETDGYVAPLTC 1139
Db      1074 PLAPSEGAGSDVFDDDLQMGAAKIGISPTHTDPSFLQYSEDFTVPLPSETDGYVAPLTC 1133
QY      1140 SPQPEYVNPQDVRPQPPSPREGPLPAAPAGATLEBAKTLSFGKNGVVKDVPFAFGAVEN 1199
Db      1134 SPQPEYVNPQDVRPQPPSPREGPLPAAPAGATLEBAKTLSFGKNGVVKDVPFAFGAVEN 1193
QY      1200 PEYLTPOGGAAPQHPHPAFSPAFPNLYYMDQDPPERGAPSTFKGTPTAENBEYGLDV 1259
Db      1194 PEYLTPOGGAAPQHPHPAFSPAFPNLYYMDQDPPERGAPSTFKGTPTAENBEYGLDV 1253
QY      1260 PV 1261
Db      1254 PV 1255

```

RESULT 11
AAE24067
ID AAE24067 standard; Protein; 1255 AA.

XX AAE24067;
XX
XX 23-SEP-2002 (first entry)
XX
XX Human Her-2 protein.
XX
XX Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
XX
XX hyperproliferative disorder; prophylaxis; inflammation; antitense;
XX
XX tumour; gene therapy; phosphotriothace backbone.
XX
XX Homo sapiens.
XX
XX WO200222636-A1.
XX

```

PD      21-MAR-2002.
XX
XX
XX 12-SEP-2001; 2001WO-US28572.
XX
XX 15-SEP-2000; 2000US-0663834.
XX
XX (ISIS-) ISIS PHARM INC.
XX
XX Bennett CF, Cowseert LM,
XX
XX WPI; 2002-471192/50.
XX
XX N-PSDB; AAD38904.
XX
XX
XX Novel antisense oligonucleotide which modulates the expression of Human
XX Epidermal Growth Factor receptor. Her2, is useful for treating tumors
XX inflammation or to prevent infection in humans -
XX
XX Example 13; Page 95-107; 116pp; English.
XX
XX The invention relates to antisense compounds targeted to a nucleic
XX acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
XX that specifically hybridises with and inhibits the expression of Her2.
XX Antisense compounds of the invention are used for treating diseases or
XX conditions associated with Her2 such as hyperproliferative disorders
XX e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
XX neural or cardiac cancer. They are also useful prophylactically e.g.
XX to prevent or delay infection, inflammation and tumour formation. The
XX invention is also used in gene therapy. The present sequence is human
XX Her-2 protein.
XX
XX Sequence 1255 AA;
XX
XX Query Match 97.1%; Score 6655; DB 23; Length 1255;
XX Best Local Similarity 97.4%; Pred. No. 0;
XX Matches 1229; Conservative 11; Mismatches 14; Indels 8; Gaps 2

```

D	534	FLRGCEVCESCRVLQGGPREYVNAHCLPCHPEQOPONGSTCGPBRADCCVCAHYKDF	593
Q	600	PCVVARCPGKVPDLSTYMPIMKFPDEEGACOPPCINCTHSCVDLDDGCPABOASPLTS	659
D	594	PFCVARCPSGVKPDLSYMPIMKFPDEEGACOPPCINCTHSCVDLDDGCPABOASPLTS	653
Q	660	IVSAVVGILLVVVGVGVGILLIKRQOKIRKYNTRILLQETELVEPLTPSGAMPNOAKR	719
D	654	IISAVVGILLVVVGVGVGILLIKRQOKIRKYNTRILLQETELVEPLTPSGAMPNOAKR	713
Q	720	ILKETELRKVKVLSGAGFTYKGIWIPDGENVYIPAIIVLEBNTSPKANKIIDEAVY	779
D	714	ILKETELRKVKVLSGAGFTYKGIWIPDGENVYIPAIIVLEBNTSPKANKIIDEAVY	773
Q	780	MAGVSPYVSRLGICLTSTVQVLTOLMPYGCILDHVENRGRLGSGDILNMCQIAKGM	839
D	774	MAGVSPYVSRLGICLTSTVQVLTOLMPYGCILDHVENRGRLGSGDILNMCQIAKGM	833
Q	840	SYLEDVRLVHRDLAARNVLSPNHVKITDFGLARLLDIDETEHADGGKVPIMKMALES	899
D	834	SYLEDVRLVHRDLAARNVLSPNHVKITDFGLARLLDIDETEHADGGKVPIMKMALES	893
Q	900	ILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPCTIDVYM	959
D	894	ILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPCTIDVYM	953
Q	960	IMVKCMWIDSECRREFELVSEFSRKARDQORFVIONEDLGPASPLDSTFYASLLEDOD	1019
D	954	IMVKCMWIDSECRREFELVSEFSRKARDQORFVIONEDLGPASPLDSTFYASLLEDOD	1013
Q	1020	MGDLVDAEEYLVPOGFFCPBPAGAGAMVHHRSSSTSGGDLTLGLEPSEBAPRS	1079
D	1014	MGDLVDAEEYLVPOGFFCPBPAGAGAMVHHRSSSTSGGDLTLGLEPSEBAPRS	1073
Q	1080	PLAPSEAGSDVPFDGLGMAKGLQSLPTHDSPLQRYSEDPVLPSETDGYVAPLTC	1139
D	1074	PLAPSEAGSDVPFDGLGMAKGLQSLPTHDSPLQRYSEDPVLPSETDGYVAPLTC	1133
Q	1140	SPOPEYVNOPDVRRQPSRPEGPLPAARPGALILEAKTILSPKXGVVQDVAFGGAEN	1199
D	1134	SPOPEYVNOPDVRRQPSRPEGPLPAARPGALILEAKTILSPKXGVVQDVAFGGAEN	1193
Q	1200	PEYLLTPQGAAPQPHPPAPSPAFDNLVYWDQPPERGAPESTFKGTPTAENPEYLLGDV	1259
D	1194	PEYLLTPQGAAPQPHPPAPSPAFDNLVYWDQPPERGAPESTFKGTPTAENPEYLLGDV	1253
Q	1260	PV 1261	
D	1254	PV 1255	

PD	21-FEB-2002.	
PF	14-AUG-2001; 2001WO-US41733.	
XX		
PR	14-AUG-2000; 2000US-225152P.	
PR	28-SEP-2000; 2000US-236428P.	
PR	21-FEB-2001; 2001US-270520P.	
XX		
PA	(CORI-) CORIXA CORP.	
PI	Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;	
PI	Mcneill PD, Vedvick TS;	
DR	WPI; 2002-280758/32.	
DR	N-PSDB; AAD32743.	
PT	Novel isolated Her-2/Neu polypeptide composition useful for therapy,	
PT	prevention and diagnosis of cancer, preferably breast cancer	
XX		
PS	Disclosure; Page 114-117; 123pp; English.	
CC	The invention relates to an isolated Her-2/Neu polypeptide composition	
CC	effective for eliciting an immune response. The invention is useful for	
CC	eliciting an immune response in a patient, where the patient is human	
CC	leukocyte antigen (HLA)-B*44 positive or is affected with breast cancer.	
CC	The composition is useful for the therapy and diagnosis of cancer,	
CC	preferably breast cancer, in pharmaceutical compositions, e.g., vaccine	
CC	and other compositions, for the diagnosis, prevention and treatment of	
CC	human malignancies, for stimulating and/or expanding T cells specific for	
CC	Her-2/Neu polypeptide and for inhibiting the development of cancer in a	
CC	patient. The invention is useful for stimulating a T cell response in a	
CC	human patient, as probe or primer for nucleic acid hybridization, to	
CC	selectively form duplex molecules with complementary stretches of the	
CC	entire Her-2/Neu gene or gene fragments of interest, to isolate a full	
CC	length gene from a suitable library, and to direct expression of a	
CC	polypeptide in appropriate host cells. The composition is useful in	
CC	prophylactic or therapeutic applications and for the treatment of cancer,	
CC	preferably for the immunotherapy of breast cancer and other Her-2/Neu-	
CC	associated malignancies. The invention is useful in gene therapy. The	
CC	present sequence is human Her-2/neu protein.	
XX		
SO	Sequence 1255 AA;	
Q	Query Match 97.1%; Score 6655; DB 23; Length 1255;	
Q	Best Local Similarity 97.4%; Pred. No. 0;	
Q	Matches 1229; Conservative 11; Mismatch 14; Indels 8; Gaps 2	
D	1 MELALCRMGLLALLPAGASNOVCTGDMKRLPASPETHLDMRHLYOGCQVVGNTL	60
Q	61 ELTYLPTNASLSFLQDIOEVOGYVLIANHQVQVPLQRLIRVGTOLFEDNALAVLNDG	120
D	61 ELTYLPTNASLSFLQDIOEVOGYVLIANHQVQVPLQRLIRVGTOLFEDNALAVLNDG	120
Q	121 DPLNNTTPVAGSPGSLRELQRLSLTEILKXGVLLQRPOLCYQDTIIMKQIFHKNOLA	180
D	121 DPLNNTTPVAGSPGSLRELQRLSLTEILKXGVLLQRPOLCYQDTIIMKQIFHKNOLA	180
Q	181 LTLIDNRSRACHPCSPMCKSGRCWSESESDQSLTRTYCAGGACRCGPLPTDCHEQC	240
D	181 LTLIDNRSRACHPCSPMCKSGRCWSESESDQSLTRTYCAGGACRCGPLPTDCHEQC	240
Q	241 AAGCTGPKASDCLACHFNHSGICELHCPALVTYVTDTPESMNPNEGRTYTGASCVTACP	300
D	241 AAGCTGPKASDCLACHFNHSGICELHCPALVTYVTDTPESMNPNEGRTYTGASCVTACP	300
Q	301 YNYLSTDVGSCTLVCEPLHNQCEVTAEDGTORCEKSKPCARVCYGGVGMNNTVAFWLR-V	359
D	301 YNYLSTDVGSCTLVCEPLHNQCEVTAEDGTORCEKSKPCARVCYGGVGMNNTVAFWLR-V	359
Q	360 PKYSASHLEEFAGCKKIFGSLAFLESEFDGDPASNTAPQYTKANSKFIGITELTGVLNIS	419
D	360 PKYSASHLEEFAGCKKIFGSLAFLESEFDGDPASNTAPQYTKANSKFIGITELTGVLNIS	419

Db	354	RAATVSANIQGFACCKKI	FOSLAF	FLPESF	DGDPASNTAF	ALQDEQ	QVPEIT	LEIT	GLVLTYS	413																																										
QY	420	AMDESLP	DSVFN	PONLOV	IGRI	LHN	GAIS	SLTL	OGIGIS	SMGLAS	RLSGGL	LHNTH	479																																							
Db	414	AMPDSLP	DSVFN	LOV	IGRI	LHN	GAIS	SLTL	OGIGIS	SMGLAS	RLSGGL	LHNTH	473																																							
QY	480	LCEVHTV	PWMQ	LFR	NHQALL	LHTAN	PEDE	CVGE	GLACHQ	LCA	RGHC	MG	GP	QVCN	CSQ	539																																				
Db	474	LCEVHTV	PWMQ	LFR	NHQALL	LHTAN	PEDE	CVGE	GLACHQ	LCA	RGHC	MG	GP	QVCN	CSQ	533																																				
QY	540	FLAGQDE	VECE	CR	LQ	L	PRE	YVNA	HC	L	PC	BE	CP	QNGS	YTC	CGP	AD	CV	CA	HY	KDP	599																														
Db	534	FLAGQDE	VECE	CR	LQ	L	PRE	YVNA	HC	L	PC	BE	CP	QNGS	YTC	CGP	AD	CV	CA	HY	KDP	593																														
QY	600	PFVCAR	CP	SV	K	ED	S	YMP	IM	K	F	P	DE	GA	C	P	C	I	N	T	H	S	C	V	D	L	D	K	G	C	P	A	E	R	A	S	P	L	T	S	659											
Db	594	PFVCAR	CP	SV	K	ED	S	YMP	IM	K	F	P	DE	GA	C	P	C	I	N	T	H	S	C	V	D	L	D	K	G	C	P	A	E	R	A	S	P	L	T	S	653											
QY	660	IVS	AV	GI	L	L	V	V	L	V	G	V	F	I	L	L	K	R	Q	O	K	I	R	Y	T	R	L	L	O	E	L	V	E	P	L	T	S	G	A	M	P	N	O	A	R	719						
Db	654	IVS	AV	GI	L	L	V	V	L	V	G	V	F	I	L	L	K	R	Q	O	K	I	R	Y	T	R	L	L	O	E	L	V	E	P	L	T	S	G	A	M	P	N	O	A	R	713						
QY	720	ILKETE	L	R	K	V	K	V	L	G	S	G	A	F	G	T	V	K	I	M	I	D	G	E	N	V	M	I	P	A	I	K	V	L	E	N	T	S	P	R	A	N	K	I	S	I	D	E	A	V	779	
Db	714	ILKETE	L	R	K	V	K	V	L	G	S	G	A	F	G	T	V	K	I	M	I	D	G	E	N	V	M	I	P	A	I	K	V	L	E	N	T	S	P	R	A	N	K	I	S	I	D	E	A	V	773	
QY	780	MAQV	S	P	Y	S	R	L	L	G	I	T	L	T	S	T	V	O	L	T	Q	L	M	P	Y	C	L	L	D	H	R	E	N	R	G	L	S	O	D	L	L	N	K	O	I	A	K	M	839			
Db	774	MAQV	S	P	Y	S	R	L	L	G	I	T	L	T	S	T	V	O	L	T	Q	L	M	P	Y	C	L	L	D	H	R	E	N	R	G	L	S	O	D	L	L	N	K	O	I	A	K	M	833			
QY	840	SYLED	R	L	V	H	R	D	L	A	R	N	V	L	K	S	P	N	H	V	K	I	T	D	F	G	L	A	R	L	D	I	D	E	T	E	Y	H	A	D	G	K	V	P	I	K	M	A	L	E	S	899
Db	834	SYLED	R	L	V	H	R	D	L	A	R	N	V	L	K	S	P	N	H	V	K	I	T	D	F	G	L	A	R	L	D	I	D	E	T	E	Y	H	A	D	G	K	V	P	I	K	M	A	L	E	S	893
QY	900	ILRR	R	T	H	S	D	S	W	S	G	V	Y	W	E	L	M	T	G	A	P	Y	C	I	T	A	R	I	P	O	L	E	K	E	R	L	P	O	P	I	C	T	I	D	V	M	959					
Db	894	ILRR	R	T	H	S	D	S	W	S	G	V	Y	W	E	L	M	T	G	A	P	Y	C	I	T	A	R	I	P	O	L	E	K	E	R	L	P	O	P	I	C	T	I	D	V	M	953					
QY	960	IMK	C	M	M	I	S</																																													

KW Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
 KM tyrosine kinase; receptor; c-erbB2; gene therapy.
 XX
 OS Homo sapiens.
 PH Key
 FT Location/Qualifiers
 FT 1..653
 FT /note="extracellular domain"
 FT 676..1255
 FT /note="intracellular domain"
 FT Domain
 FT 990..1255
 FT /note="phosphorylation domain"
 FT
 PN WO200212341-A2.
 PD 14-FEB-2002.
 XX
 XX 03-AUG-2001; 2001WO-US24283.
 PF 03-AUG-2000; 2000US-0632507.
 PR 03-AUG-2000; 2000US-0632507.
 XX
 XX (CORI-) CORIXA CORP.
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX
 XX Cheever MA, Ghesen D;
 PL
 DR WPJ; 2002-241743/29.
 DR N-PSDB; ABA922250.
 XX
 PT Her-2/neu fusion protein for treating or preventing cancer by eliciting
 PT or enhancing an immune response to the protein, has Her-2/neu
 PT extracellular domain fused to Her-2/neu intracellular or
 PT phosphorylation domain -
 XX
 XX
 PS Claim 68; Fig 7; 14pp; English.
 XX
 CC The present sequence is that of human Her-2/neu (p185 glycoprotein
 CC or c-erbB2) an oncogenic self-protein and target for anti-cancer
 CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
 CC in a variety of cancers, including breast, ovarian, colon, lung and
 CC prostate cancer. Her-2/neu is a member of the tyrosine kinase
 CC family of receptor-like glycoproteins. It comprises an extracellular
 CC domain with homology to the epidermal growth factor receptor
 CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
 CC intracellular domain that also shows homology to EGFR. Its
 CC overexpression correlates with a poor prognosis in breast and
 CC ovarian cancers. The invention provides Her-2/neu fusion
 CC proteins, nucleic acids encoding them, viral vectors, and vaccines
 CC comprising the fusion proteins or nucleic acid molecules. In
 CC preferred fusion proteins, the extracellular domain of a Her-2/neu
 CC protein is fused to a Her-2/neu intracellular domain or
 CC phosphorylation domain (or its DeltaPD fragment). An immune
 CC response to Her-2/neu protein is elicited or enhanced by
 CC administering the fusion protein in the form of a vaccine, or by
 CC transfecting cells of an animal *ex vivo* with a nucleic acid
 CC encoding the fusion protein, and delivering the transfected cells
 CC to the animal. The fusion proteins, nucleic acids, and isolated
 CC specific T-cells are useful for inhibiting the development of a
 CC cancer, especially breast, ovarian, colon, lung or prostate cancer
 CC in a patient. T cells that specifically react with a Her-2/neu
 CC fusion protein can be used to remove tumour cells from a sample in
 CC order to inhibit the development of cancer in a patient.
 CC
 SO Sequence 1255 AA;
 OY
 Query Match 97.1%; Score 6655; DB 23; Length 1255;
 Best Local Similarity 97.4%; Pred. No. 0;
 Matches 1229; Conservative 11; Mismatches 14; Indels 8; Gaps 2.
 1 METALACWGLLALLPPGASATVCGCTGTMKRLPASPEPTHLDMLRLHYOCQVQVGLT 60
 1 METALACWGLLALLPPGASATVCGCTGTMKRLPASPEPTHLDMLRLHYOCQVQVGLT 60

```

QY 61 ELTYLPFTNASLSFLODIOEVGVYLIANOVQVPLQRLIRIVRGTOLEFEDNYALAVLDNG 120
DB 61 ELTYLPFTNASLSFLODIOEVGVYLIANOVQVPLQRLIRIVRGTOLEFEDNYALAVLDNG 120
QY 121 DPLNNTPTVGTASGGGLRELOLRSLTELLKGVYLIQNPOLCQODITLMKDIHKNNOA 180
DB 121 DPLNNTPTVGTASGGGLRELOLRSLTELLKGVYLIQNPOLCQODITLMKDIHKNNOA 180
QY 181 LTLIDITNRSRACHPCSPMCKGSRCMGSSSDCQSLTRTVACAGGACAKGPLPTDCCHQC 240
DB 181 LTLIDITNRSRACHPCSPMCKGSRCMGSSSDCQSLTRTVACAGGACAKGPLPTDCCHQC 240
QY 241 AAGCTGPGHSDCLCLPHNHSIGICELHCPALVYNTDTFESMNPBGRYTFGASCTTAC 300
DB 241 AAGCTGPGHSDCLCLPHNHSIGICELHCPALVYNTDTFESMNPBGRYTFGASCTTAC 300
QY 301 YNYLSTDVGSCTLVCPHNOEVTAEEDGTORCEKSKPCAVCYGLMFPNFTVSFWLR-V 359
DB 301 YNYLSTDVGSCTLVCPHNOEVTAEEDGTORCEKSKPCAVCYGLMFPNFTVSFWLR-V 359
QY 360 PKVSASHLEEFAGCKKTFGSLAF-PESFGDDPASNTAPQYIKANSKFGITELTGYLYIS 419
DB 360 PKVSASHLEEFAGCKKTFGSLAF-PESFGDDPASNTAPQYIKANSKFGITELTGYLYIS 419
QY 420 AMPDLPDLISYFQNLQYIRGRILHNGAYSLTLOGLISWLGRLSRLSGLALIHNNH 479
DB 420 AMPDLPDLISYFQNLQYIRGRILHNGAYSLTLOGLISWLGRLSRLSGLALIHNNH 479
QY 480 LCPVHTVPMQDLFNNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPPTOCVNC 539
DB 480 LCPVHTVPMQDLFNNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPPTOCVNC 539
QY 540 FLRGQCEVEECRVIOGLPREYVNAHCLPCHPECOPONGSVTCFGEANOCVACAKYKP 599
DB 540 FLRGQCEVEECRVIOGLPREYVNAHCLPCHPECOPONGSVTCFGEANOCVACAKYKP 599
QY 594 PFCVAPCPSPGVKPLSLTMRPKFPEDEGACQPCPINCHTSCVDLDKGPACRASPILTS 653
DB 594 PFCVAPCPSPGVKPLSLTMRPKFPEDEGACQPCPINCHTSCVDLDKGPACRASPILTS 653
QY 660 IVSAVVGILLVYLVGVFGLILKRQOKIRKYMRLLOETELVEPLTPSGAMPQAOQR 719
DB 660 IVSAVVGILLVYLVGVFGLILKRQOKIRKYMRLLOETELVEPLTPSGAMPQAOQR 719
QY 720 ILKETELRKVYVGSAGFTVYKGIWIPDGENVKITVAIKVLRNENSPANKELIDEAV 779
DB 720 ILKETELRKVYVGSAGFTVYKGIWIPDGENVKITVAIKVLRNENSPANKELIDEAV 779
QY 780 MAGVGSPIYVSRLLIGICTLSTVQLVTQIMPYGCLLDHVRENRGLSGODLLNMCQIAK 839
DB 780 MAGVGSPIYVSRLLIGICTLSTVQLVTQIMPYGCLLDHVRENRGLSGODLLNMCQIAK 839
QY 840 SYLEDVRLVHRDLAARVAVLKSPNHVKITDFGLARLLDIDEVHADGGKVIKMALES 899
DB 840 SYLEDVRLVHRDLAARVAVLKSPNHVKITDFGLARLLDIDEVHADGGKVIKMALES 899
QY 900 ILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI-PAREIDLEKEBRLLPQPPICIDVYM 959
DB 900 ILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI-PAREIDLEKEBRLLPQPPICIDVYM 959
QY 960 IMYKCMWIDSECRPRELVESESRMAKPPQRPVVIQNDLGPASPLSTFPRSLLEDD 1019
DB 960 IMYKCMWIDSECRPRELVESESRMAKPPQRPVVIQNDLGPASPLSTFPRSLLEDD 1019
QY 1020 MGDLDVAEEYLVPOQGFPCDPAPAGAGWVHRHRSSTRSGGDLTLGLESESEAPRS 1079
DB 1020 MGDLDVAEEYLVPOQGFPCDPAPAGAGWVHRHRSSTRSGGDLTLGLESESEAPRS 1079
QY 1080 PLAPSEGAGSDVDGDLGMAKGLQSLTPHPSPLQRYSEDPYVPLPSETIDGYAPLTC 1139
DB 1080 PLAPSEGAGSDVDGDLGMAKGLQSLTPHPSPLQRYSEDPYVPLPSETIDGYAPLTC 1139
QY 1140 SPOPEYVNOFDVPRQPPSPREGPLPAARPAAGATLEBAKTLSPKNGVYKDVFAFGA 1199
DB 1140 SPOPEYVNOFDVPRQPPSPREGPLPAARPAAGATLEBAKTLSPKNGVYKDVFAFGA 1199

```

```

DB 1134 SPOPEYVNOFDVPRQPPSPREGPLPAARPAAGATLEBAKTLSPKNGVYKDVFAFGA 1199
QY 1200 PEYLTPQGGAAPQCHPPAPASPAFDNLYYDODPPEGAPSPSTFKGTPTANEYGLDV 125
DB 1194 PEYLTPQGGAAPQCHPPAPASPAFDNLYYDODPPEGAPSPSTFKGTPTANEYGLDV 125
QY 1260 PV 1261
DB 1254 PV 1255

RESULT 14
AAU77114
ID AAU77114 standard; Protein; 1255 AA.
AC AAU77114;
XX 05-JUN-2002 (first entry)
DT
DE Human Her-2/neu polypeptide.
KW Human; Her-2/neu; cytostatic; hematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
OS Homo sapiens.
PN MO200213647-A2.
XX 21-FEB-2002.
PD
XX 13-AUG-2001; 2001MO-US25408.
PE
XX 14-AUG-2000; 2000JUS-0638280.
PR 28-SEP-2000; 2000JUS-0675904.
XX
PA (CORI-) CORIXA CORP.
XX
PI Gaiger A, Cheever MA, Hand-zimmermann S;
XX WPI: 2002-280741/32.
DR N-PSDB; ABK10730.
XX
FT Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide
XX
PS Disclosure; Page 71-74; 74pp; English.
XX
CC The invention relates to a method for inhibiting development of
CC hematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of hematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;
Query Match 97.1%; Score 6655; DB 23; Length 1255;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1229; Conservative 11; Mismatches 14; Indels 8; Gaps 2;
QY 1 METALACRGGILLALLPFGAAGTQVCTGDKMLRLPASPETHLMDLRLYOGCGVQGNL 60
DB 1 METALACRGGILLALLPFGAAGTQVCTGDKMLRLPASPETHLMDLRLYOGCGVQGNL 60
QY 61 ELTYLPFTNASLSFLODIOEVGVYLIANOVQVPLQRLIRIVRGTOLEFEDNYALAVLDNG 120

```

```

Db      61 ELTYLPNTASLSFLDIQEOGYVLIANQVRQVPLQRRLRVKRGTOQLFENNYALAVLDNG 120
Qy      121 DPLNNTPTVTGASPGGLRELOLRSLTEILKGVLLIQRNPOLCYODTILMKDIFHKNNOLA 180
Db      121 DPLNNTPTVTGASPGGLRELOLRSLTEILKGVLLIQRNPOLCYODTILMKDIFHKNNOLA 180
Qy      181 LTLIDNRSRACHPSPCKSGSRCKWGESSEDCOSLTRTVACGAGARCKGPLPTDCCHQC 240
Db      181 LTLIDNRSRACHPSPCKSGSRCKWGESSEDCOSLTRTVACGAGARCKGPLPTDCCHQC 240
Qy      241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNBSGRYTFGASCTYACP 300
Db      241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNBSGRYTFGASCTYACP 300
Qy      301 YNYLSTDVSGCTVCPILHNOEVTAEQDQRCCKSKCAVCVGLGFNNFTYSFMLR-V 359
Db      301 YNYLSTDVSGCTVCPILHNOEVTAEQDQRCCKSKCAVCVGLGFNNFTYSFMLR-V 359
Qy      360 PKVSASHLEEFAGCKKIFGSLAPLPESHDGDPASNTAPQYIKANSKFIGITETGYLIS 419
Db      354 RAVTSANIQEFAGCKKIFGSLAPLPESHDGDPASNTAPLPQEQQLVETLEETLGTLYIS 413
Qy      420 AMPDSLPLDSVFQNLQVIRGRIHNGAYSLTLQIGISWIGLSLRGLSGLALIHNTH 479
Db      414 AMPDSLPLDSVFQNLQVIRGRIHNGAYSLTLQIGISWIGLSLRGLSGLALIHNTH 473
Qy      480 LCFVHTVPMQDLFRNPQALHTANRPEDECEVGEGLAQHQLCARGHGCMWGPPTQCVNCSO 539
Db      474 LCFVHTVPMQDLFRNPQALHTANRPEDECEVGEGLAQHQLCARGHGCMWGPPTQCVNCSO 533
Qy      540 FLRGOECVECRVYIGLPREYVNAHCLPCHPECOPONGSVTCGPBACDVACAAYKDP 599
Db      534 FLRGOECVECRVYIGLPREYVNAHCLPCHPECOPONGSVTCGPBACDVACAAYKDP 593
Qy      600 PFCVACRCPGVKPDLSYMPIMKFPDEBGAOCPCINTHSCVLDKDGCAEBSAPLS 659
Db      594 PFCVACRCPGVKPDLSYMPIMKFPDEBGAOCPCINTHSCVLDKDGCAEBSAPLS 653
Qy      660 TVSAVVGILLVVLGVVFGILLIKRQCKIRKYTNRLLOETELVEPLTPSGAMPNQAQNR 719
Db      654 TVSAVVGILLVVLGVVFGILLIKRQCKIRKYTNRLLOETELVEPLTPSGAMPNQAQNR 713
Qy      720 ILKETELRKXVULSGAGTGYKGIWIPDGENYKIPAIKVLBNENSPKANKILDEAVY 779
Db      714 ILKETELRKXVULSGAGTGYKGIWIPDGENYKIPAIKVLBNENSPKANKILDEAVY 773
Qy      780 MAGVSSPYVSRLLGICLTSTVQLVTQLMPYGCILDHYRENRGRIGSODLLNMCQIAKGM 839
Db      774 MAGVSSPYVSRLLGICLTSTVQLVTQLMPYGCILDHYRENRGRIGSODLLNMCQIAKGM 833
Qy      840 SYLEDVRLVHARDLAARNVLVKSPNHVKITDGLARLIDIDETEHADGKVPKIMALES 899
Db      834 SYLEDVRLVHARDLAARNVLVKSPNHVKITDGLARLIDIDETEHADGKVPKIMALES 893
Qy      900 ILRRRFTHQSDVMSGYVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPORPCTIDVYM 959
Db      894 ILRRRFTHQSDVMSGYVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPORPCTIDVYM 953
Qy      960 IMVKCMIMIDSECRPFEBELVSEFSRMARDPQRFVYIIONEDLGASPLDSTFYRSLLEDD 1019
Db      954 IMVKCMIMIDSECRPFEBELVSEFSRMARDPQRFVYIIONEDLGASPLDSTFYRSLLEDD 1013
Qy      1020 MGDLDVAEEYLVPOGGFFCPDPAPAGAGMHHNRSSSTRSGGDLTLGLEBEEEPAPS 1079
Db      1014 MGDLDVAEEYLVPOGGFFCPDPAPAGAGMHHNRSSSTRSGGDLTLGLEBEEEPAPS 1073
Qy      1080 FLAPSEGAGSVEFDGDIQKGAAGKLGSLPHDPSPLQRYSEDEPTVPLPSETDGYVAPLTC 1139
Db      1074 FLAPSEGAGSVEFDGDIQKGAAGKLGSLPHDPSPLQRYSEDEPTVPLPSETDGYVAPLTC 1133
Qy      1140 SPOPEYVNUQPVRRQPPSPREGPLPARPAGATLERAKTILSPKNGVVDVFAFGAVEN 1199

```

```

Db      1134 SPOPEYVNUQPVRRQPPSPREGPLPARPAGATLERAKTILSPKNGVVDVFAFGAVEN 119
Qy      1200 PEYLTPOGGAAPQHPHPAPSPAFDNLVYWDQDPERGAPSTFFKGTPTAENDEYVGLDV 125
Db      1194 PEYLTPOGGAAPQHPHPAPSPAFDNLVYWDQDPERGAPSTFFKGTPTAENDEYVGLDV 125
Qy      1260 PV 1261
Db      1254 PV 1255

RESULT 15
AAR39568
ID AAR39568 standard; Protein: 1433 AA.
XX
AC AAR39568;
XX
DT 07-FEB-1994 (first entry)
XX
DE Sequence of c-erbB-2 tumour antigen.
XX
KW Tumour antigen; c-erbB-2; glycoprotein.
XX
OS Homo sapiens.
XX
PN W09316185-A.
XX
PD 19-AUG-1993.
XX
PF 05-FEB-1993; 93WC-US01055.
XX
PR 06-FEB-1992; 92US-0831967.
XX
PA (CETU ) CETUS ONCOLOGY CORP.
XX
PA (CREA-) CREATIVE BIOMOLECULES INC.
XX
PI Houston IL, Huston JS, Opfermann H, Ring DB;
XX
DR WPI; 1993-272889/34.
XX
DR N-PEDB; AAQ46083.
XX
PT New single chain Fv polypeptide binding to C-erbB-2 tumour
XX
PT antigen - for imaging or treating breast or ovarian cancer etc.
XX
PS Disclousure; pages 48-54; 87pp; English.
XX
CC c-erbB-2 refers to a protein antigen expressed on the surface of
XX
CC tumour cells such as breast and ovarian tumour cells, which is an
XX
CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
XX
CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39568 represents
XX
CC the location of a stop codon in AAQ46083.
XX
SQ Sequence 1433 AA;

Query Match 96.5%; Score 6616; DB 14; Length 1433;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 1223; Conservative 13; Mismatches 18; Indels 8; Gaps 2

Qy      1 MELALCRMGILLALLPPGAASQVCTGTDMKRLPASDETHLDMRLHYOGQOVVQGNL 60
Db      1 MELALCRMGILLALLPPGAASQVCTGTDMKRLPASDETHLDMRLHYOGQOVVQGNL 60
Qy      61 ELTYLPNTASLSFLDIQEOGYVLIANQVRQVPLQRRLRVKRGTOQLFENNYALAVLDNG 120
Db      61 ELTYLPNTASLSFLDIQEOGYVLIANQVRQVPLQRRLRVKRGTOQLFENNYALAVLDNG 120
Qy      121 DPLNNTPTVTGASPGGLRELOLRSLTEILKGVLLIQRNPOLCYODTILMKDIFHKNNOLA 180
Db      121 DPLNNTPTVTGASPGGLRELOLRSLTEILKGVLLIQRNPOLCYODTILMKDIFHKNNOLA 180
Qy      181 LTLIDNRSRACHPSPCKSGSRCKWGESSEDCOSLTRTVACGAGARCKGPLPTDCCHQC 240
Db      181 LTLIDNRSRACHPSPCKSGSRCKWGESSEDCOSLTRTVACGAGARCKGPLPTDCCHQC 240

```

Search completed: July 22, 2003, 08:41:29
Job time : 44.1589 secs

```
QY 241 AAGCTGPHSDCLACLFHNSGICELHCPALVTYNTCTFESMNPBGRYTFGASCYTACP 300
DB 241 AAGCTGPHSDCLACLFHNSGICELHCPALVTYNTCTFESMNPBGRYTFGASCYTACP 300
QY 301 YNYLSTDVSGCTLYCPHNOEVTAEADGTQCEKSKPCARVCYGLGMFNNFTVFWLR-V 359
DB 301 YNYLSTDVSGCTLYCPHNOEVTAEADGTQCEKSKPCARVCYGLGMFNNFTVFWLR-V 359
QY 360 PKVSASHLEEFAGCKTFGSLAFLPESFGDDPASNTAPQYIKANSKFITELTGLYIS 419
DB 360 PKVSASHLEEFAGCKTFGSLAFLPESFGDDPASNTAPQYIKANSKFITELTGLYIS 419
QY 420 AMPDLSPLDSVFQNLQYIRGRILHNGAVSLTLOGLISWGLSLRELGSGLLIHNTH 479
DB 420 AMPDLSPLDSVFQNLQYIRGRILHNGAVSLTLOGLISWGLSLRELGSGLLIHNTH 479
QY 474 LCFVHTVPMDOLEFNPQALLHTANRPEDCEVGEGLACHOLCARGHCMGPPTQCVNCSQ 539
DB 474 LCFVHTVPMDOLEFNPQALLHTANRPEDCEVGEGLACHOLCARGHCMGPPTQCVNCSQ 539
QY 540 FLRGQECVEECRVQLGPREVNAHCLPCHPECPONGSVTCFGEADQVCAHAYKDP 599
DB 540 FLRGQECVEECRVQLGPREVNAHCLPCHPECPONGSVTCFGEADQVCAHAYKDP 599
QY 594 PFCVARCPSGVKPLSYMPIMKFPDEEGACQPCPINCTHSCVDLDDKGCPRAEQRASPLTS 659
DB 594 PFCVARCPSGVKPLSYMPIMKFPDEEGACQPCPINCTHSCVDLDDKGCPRAEQRASPLTS 659
QY 660 IVSAVVGILLVVLGVVFGILIKRROKIRKYMRLLOETELVEBLTPSGAMPNQAQMR 719
DB 660 IVSAVVGILLVVLGVVFGILIKRROKIRKYMRLLOETELVEBLTPSGAMPNQAQMR 719
QY 720 ILKETELRKVKVLGSGAFGTVYKGIWIPGSENVKIPVAIKVLRENTSPRANKELLDEAYV 779
DB 720 ILKETELRKVKVLGSGAFGTVYKGIWIPGSENVKIPVAIKVLRENTSPRANKELLDEAYV 779
QY 774 MAGVSPYVSRLLGICLTSTVOLVTQLMRYGCLLDVYRENRGLSGQDLLNMCMQIAKGM 839
DB 774 MAGVSPYVSRLLGICLTSTVOLVTQLMRYGCLLDVYRENRGLSGQDLLNMCMQIAKGM 839
QY 840 SYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLJIDETEHADGGKVPIMXMALES 899
DB 840 SYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLJIDETEHADGGKVPIMXMALES 899
QY 900 ILRRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICIIDVYM 959
DB 900 ILRRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICIIDVYM 959
QY 960 IMYKCMWIDSECRPRFELVSEFSRMAKRDQRFVVIQNEDLGPASPLDSTFYRSLIEDDD 1019
DB 960 IMYKCMWIDSECRPRFELVSEFSRMAKRDQRFVVIQNEDLGPASPLDSTFYRSLIEDDD 1019
QY 1020 MGDLYDAEEXLVLPQGFPCDPAPAGAGMTHHRHRSSTRSGGDLTLGLPESEBEAPRS 1079
DB 1020 MGDLYDAEEXLVLPQGFPCDPAPAGAGMTHHRHRSSTRSGGDLTLGLPESEBEAPRS 1079
QY 1080 PLAPSEAGSDVDVDGDLGMAAGKLOSLPTHDSPLORXSEDPVPLPSETDGYVAPLTC 1139
DB 1080 PLAPSEAGSDVDVDGDLGMAAGKLOSLPTHDSPLORXSEDPVPLPSETDGYVAPLTC 1139
QY 1140 SPOEYVNOQDVVRPOPSPREGFLPARAPAGATLERAKTLSPGKNGVAVKDVAFGAVEN 1199
DB 1140 SPOEYVNOQDVVRPOPSPREGFLPARAPAGATLERAKTLSPGKNGVAVKDVAFGAVEN 1199
QY 1200 PEYLTTPQGAAPQHPPAPSPAFDNLVYWDODPBERGAPSTFKGTPTAENPEYLGLDV 1259
DB 1200 PEYLTTPQGAAPQHPPAPSPAFDNLVYWDODPBERGAPSTFKGTPTAENPEYLGLDV 1259
QY 1260 PV 1261
DB 1254 PV 1255
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.1815 Seconds

(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-5-25-12

Perfect score: 6774

Sequence: 1 MELALCRWGLLALPPGA.....TFKGTPTAENPEYLGDVY 1249

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6672	98.5	1255	1	ERB2_HUMAN
2	5861.5	86.5	1254	1	ERB2_MESAU
3	5860	86.5	1257	1	ERB2_RAT
4	3133	46.3	1210	1	EGFR_HUMAN
5	3115	46.0	1210	1	EGFR_MOUSE
6	2968.5	43.8	1308	1	ERB4_HUMAN
7	2949	43.5	1308	1	ERB4_RAT
8	2696.5	39.8	1167	1	XMRK_XIPMA
9	2423	35.8	1142	1	ERB3_HUMAN
10	2356.5	34.8	1339	1	EGFR_DROME
11	1927	28.4	1426	1	EGFR_MOUSE
12	1749.5	25.8	1434	1	ERBB_DROME
13	1703	25.1	604	1	ERBB_AVIER
14	1630	24.1	540	1	ERBB_XIPMA
15	1586	23.4	703	1	EGFR_CHICK
16	1300.5	19.2	1323	1	L1232_GAEL
17	1142.5	16.9	245	1	ERB2_MOUSE
18	722	10.7	1363	1	ILPR_BRALA
19	698	10.3	1300	1	IRR_MOUSE
20	696.5	10.3	1382	1	INSR_HUMAN
21	694.5	10.3	1607	1	WIPR_LYST
22	691	10.2	1372	1	INSR_MOUSE
23	690	10.2	1297	1	IRR_HUMAN
24	687.5	10.1	1477	1	INSR_RAT
25	687.5	10.1	1477	1	HTK7_HYDAT
26	683.5	10.1	1300	1	IRR_GAVPO
27	654	9.7	1367	1	IGIR_HUMAN
28	644	9.5	1373	1	IGIR_MOUSE
29	640.5	9.5	1370	1	IGIR_RAT
30	612.5	9.0	2146	1	INSR_DROME
31	610.5	9.0	1390	1	INSR_AEDAE
32	603	8.9	987	1	EPB4_HUMAN
33	591.5	8.7	984	1	EPB1_CHICK

34	589.5	8.7	977	1	EPB2_MOUSE	C03145 mus musculus
35	588	8.7	1114	1	RET_HUMAN	P07949 homo sapien
36	584.5	8.6	976	1	EPB3_HUMAN	P29317 homo sapien
37	583.5	8.6	984	1	EPB1_RAT	P09759 rattus norv
38	582	8.6	987	1	EPB4_MOUSE	P24761 mus musculus
39	579	8.5	902	1	EPB8_XENLA	Q91736 xenopus lae
40	577.5	8.5	984	1	EPB1_HUMAN	P24762 homo sapien
41	576.5	8.5	985	1	EPB4_XENLA	Q91571 xenopus lae
42	573.5	8.5	1053	1	FAK1_CHICK	O00944 gallus gall
43	569	8.4	1068	1	FAK1_XENLA	Q91738 xenopus lae
44	567	8.4	757	1	HTK7_HYDAT	P33356 hydra atten
45	563	8.3	1052	1	FAK1_MOUSE	P34152 mus musculus

ALIGNMENTS

RESULT 1
ID ERB2_HUMAN STANDARD; PRT; 1255 AA.
AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p155erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE surface receptor HER2) (MLN 15').
GN ERB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86118663; PubMed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Senda K., Nomura N., Miyajima N.,
RA Saito T., Toyoshima K.,
RT "Similarity of protein encoded by the human c-erbB-2 gene to
RT epidermal growth factor receptor.";
RL Nature 319:230-234(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86070181; PubMed=2999974;
RA Cousens L., Yang-Feng T.L., Liao Y.C., Chen F., Gray A.,
RA McGrath U., Seeburg P.H., Ullrich A., Schlessinger J.,
RA Francke U., Levinson A., Ullrich A.,
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene.";
RL Science 230:1132-1139(1985).
RN [3]
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=86016729; PubMed=295967;
RA Senda K., Kanata N., Toyoshima K., Yamamoto T.,
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor receptor gene and is amplified in a
RT human salivary gland adenocarcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN [4]
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.,
RT "Characterization of a new allele of the human ERBB2 gene by allele-
RT specific competition hybridization.";
RL Genomics 15:426-429(1993)
CC -I- FUNCTION. ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -I- CATALYTIC ACTIVITY. ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -I- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIALLY).
CC -I- SUBCELLULAR LOCATION: Type I membrane protein.

[illegible]

```

QY 715 RKVKLGSGAFGVYKGIWIPDGENVAKIPVAIKVLEBENTSPKANKEILDEAYVMGVGSP 774
DB 721 RKVKLGSGAFGVYKGIWIPDGENVAKIPVAIKVLEBENTSPKANKEILDEAYVMGVGSP 780
QY 775 YVSRLLGICLTSTVQVLTOLMPYGCILDYHRENRKGLSGODLLNWMQIAKMSYLEDR 834
DB 781 YVSRLLGICLTSTVQVLTOLMPYGCILDYHRENRKGLSGODLLNWMQIAKMSYLEDR 840
QY 835 LVHRDLAARNVLKSPNVAKITDFGLARLIDIDETEVYHADGKGVPIKMWALESIIRRRPT 894
DB 841 LVHRDLAARNVLKSPNVAKITDFGLARLIDIDETEVYHADGKGVPIKMWALESIIRRRPT 900
QY 895 HQSDVMSYGVTVWELMTFGAKXPFDGIPARIPDLLEKGERLPOPPICITDVTMIMWKKMM 954
DB 901 HQSDVMSYGVTVWELMTFGAKXPFDGIPARIPDLLEKGERLPOPPICITDVTMIMWKKMM 960
QY 955 IDSECRPRFRELSEFMRMARDQRFVYVIONEDLGPASPLDSTFVYSILEDDMDGLVDA 1014
DB 961 IDSECRPRFRELSEFMRMARDQRFVYVIONEDLGPASPLDSTFVYSILEDDMDGLVDA 1020
QY 1015 EBYLVPOQGFCCDPAPAGAGVYHHRSSSTRSGGGDTTLGLEPSEEBAPRSPPLAPSEGG 1074
DB 1021 EBYLVPOQGFCCDPAPAGAGVYHHRSSSTRSGGGDTTLGLEPSEEBAPRSPPLAPSEGG 1080
QY 1075 AGSDVDFGDLGMAAGKGLSLPTHDSPLOKVSSEDPVTLBSETOGYAPLTCSPQPEYV 1134
DB 1081 AGSDVDFGDLGMAAGKGLSLPTHDSPLOKVSSEDPVTLBSETOGYAPLTCSPQPEYV 1140
QY 1135 NOPDVRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVDFVAFGAVENPEYLTPO 1194
DB 1141 NOPDVRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVDFVAFGAVENPEYLTPO 1200
QY 1195 GGAAPQHPPPAPSPAFDNLXYWDDPPERGAPPSFKTKITPAENPEYIGLDVPIV 1249
DB 1201 GGAAPQHPPPAPSPAFDNLXYWDDPPERGAPPSFKTKITPAENPEYIGLDVPIV 1255

RESULT 2
ERB2_MESAU STANDARD; PRT; 1254 AA.
ID ERB2_MESAU STANDARD; PRT; 1254 AA.
AC Q60353;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 37, Last sequence update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
CN ERBB2 OR NEU
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxId=10036;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94193007; PubMed=7908275;
RA Nakamura T., Ishikawa T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255(1994).
CC -|- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULIN DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -|- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -|- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -|- SUBCELLULAR LOCATION: Type I membrane protein.
CC -|- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -|- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

```

```

CC CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D16295; BAA03801.1; -
CC HSP: P11363; 1FCX
CC InterPro: IPR000494; EGFR_L_domain.
CC InterPro: IPR000719; Euk_kinase.
CC InterPro: IPR002174; Furin-like.
CC InterPro: IPR001245; Tyr_kinase.
CC InterPro: IPR004019; YLP motif.
CC Pfam: PF00069; kinase_1.
CC Pfam: PF00757; Furin-like; 1.
CC Pfam: PF01030; Recep_L_domain; 2.
CC Pfam: PF02757; YLP; 2.
CC ProDom: PD000001; Euk_kinase; 1.
CC SMART: SM00261; Fu; 3.
CC SMART: SM00219; Tyrc; 1.
CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE: PS00109; PROTEIN_KINASE_TIR; 1.
CC PROSITE: PS0011; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
CC Proto-oncogene; Disease mutation.
CC KX SIGNAL 1 21
CC FT CHAIN 1 21
CC FT DOMAIN 22 1254
CC FT TRANSMEM 22 652
CC FT DOMAIN 653 675
CC FT DOMAIN 676 1254
CC FT DOMAIN 158 368
CC FT DOMAIN 472 644
CC FT DOMAIN 720 987
CC FT NP_BIND 726 734
CC FT BINDING 753 753
CC FT ACT_SITE 845 845
CC FT DISULFID 195 204
CC FT DISULFID 199 212
CC FT DISULFID 236 244
CC FT DISULFID 240 252
CC FT DISULFID 255 264
CC FT DISULFID 268 295
CC FT DISULFID 299 311
CC FT DISULFID 315 331
CC FT DISULFID 334 338
CC FT DISULFID 511 520
CC FT DISULFID 515 528
CC FT DISULFID 531 540
CC FT DISULFID 544 560
CC FT DISULFID 563 576
CC FT DISULFID 584 584
CC FT DISULFID 587 596
CC FT DISULFID 600 623
CC FT DISULFID 626 634
CC FT DISULFID 630 642
CC FT MOD_RES 1139
CC FT MOD_RES 1247
CC FT CARBOHYD 68 68
CC FT CARBOHYD 125 125
CC FT CARBOHYD 187 187
CC FT CARBOHYD 259 259
CC FT CARBOHYD 330 330
CC FT CARBOHYD 571 571
CC FT CARBOHYD 629 629
CC FT VARIANT 658 658
CC FT VARIANT 659 659
CC FT SEQUENCE 1254 AA; 138252 MW; 974C3791C2A1F2B1 CR664;

```

Query Match 86.5%; Score 5861.5; DB 1; Length 1254;

Best Local Similarity 86.1%; Pred. No. 76-305;
Matches 1081; Conservative 59; Mismatches 108; Indels 7; Gaps 2;

```

QY 1 MELALACWGLLLALLPFGAASSTOVCT-----QYIANKSKFGITFELLYQGGCVTVQNL 54
DB 1 MELAMCGWGLLLALLPFGASGTCTGCTDMKLRLPASPTHTLDIRHLYQGGCVTVQNL 60
QY 55 ELTLPNTASISFODIOEVQGYLLAHNOVRQYPLRLIRVGTQLFEDNYALAVLDNG 114
DB 61 ELTLPANATLSFODIOEVQGYLLAHNOVRHPLRLIRVGTQLFEDNYALAVLDNR 120
QY 115 DPLNNTPTVTSASGGIRELQRLSLTILKGGVLIQNPOLCYQDTILWKDLFHKNOQA 174
DB 121 DPLNNTATATGRTPEGIRELQRLSLTILKGGVLIQNPOLCYQDTIVLWMDVPRKKNQA 180
QY 175 LTLIDITNSRACHPCSPKCGSRCMGSSSEDCOSLRTFTVAGGACAKCPKLPDCCHEQC 234
DB 181 PVDIDITNSRACHPCACPKCKDNHCMGASPEDCQTLITITIPRAVPAARALPTDCCHEQC 240
QY 235 AAGCTGPRHSDCLALCFHNSGICELHCPALVTYNTDTFESMNPREGRYTFGASCVTACP 294
DB 241 AAGCTGPRHSDCLALCFHNSGICELHCPALVTYNTDTFESMNPREGRYTFGASCVTACP 300
QY 295 YNYLSTDVSGCTLWCPHNOEVTAEQTCCKSKCAVCYGLGMEHLRGARALTAN 354
DB 301 YNYLSTEVSGCTLWCPHNOEVTAEQTCCKSKCAVCYGLGMEHLRGARALTAN 360
QY 355 IOEFAGCKKIFGSLAFLESPFDGPASNTAPLOEQLOVELEETLGYLISAMPDULP 414
DB 361 IOEFAGCKKIFGSLAFLESPFDGNSSGIALPTEQLOVELEETLGYLISAMPDULH 420
QY 415 DLSVFQNLQVIRGILINNGAYSLTLOGLGISWGLRLRLSGGLALIHNTHLCTVHTV 474
DB 421 DLSVFQNLQVIRGILINNGAYSLTLOGLGISWGLRLRLSGGLALIHNTHLCTVHTV 480
QY 475 PMDQLEFNPHQALHTANRPEDECVCEGLACHOLCARGHCMGSPGTCVNCOSQFLRGQC 534
DB 481 PMDQLEFNPHQALHTANRPEDECVCEGLACHOLCARGHCMGSPGTCVNCOSQFLRGQC 540
QY 535 VEECRVLOGLFREVYNARHCLPCHPECOQONGSVTCGFPADOCVAACAHKDPFCVAAC 594
DB 541 VEECRVLOGLFREVYNARHCLPCHPECOQONGSVTCGFPADOCVAACAHKDPFCVAAC 600
QY 595 PSQVPRDLSYMPIKFPDEBEGACPCINCTHSCVDLDDMGCAEGRASPLTISIVAVNG 654
DB 601 PSQVPRDLSYMPIKFPDEBEGACPCINCTHSCVDLDDMGCAEGRASPLTISIVAVNG 660
QY 655 ILVAVLGVVFGILIKRQOKIRKXKTRMLLOETELVEPLTPSGAMPNOAKRILKETEL 714
DB 661 ILVAVLGVVFGILIKRQOKIRKXKTRMLLOETELVEPLTPSGAMPNOAKRILKETEL 720
QY 715 RKVAVLGSAGATYYKGIWIPDGENVKIPVAIVLRENTSPKANKETILDEAYVMAGVSP 774
DB 721 RKVAVLGSAGATYYKGIWIPDGENVKIPVAIVLRENTSPKANKETILDEAYVMAGVSP 780
QY 775 YVSRLLGICLSTIVQLTQMLPYGCLLDHVRKGRGLSGDILLNMCVQAKGMSYLEDVR 834
DB 781 YVSRLLGICLSTIVQLTQMLPYGCLLDHVRKGRGLSGDILLNMCVQAKGMSYLEDVR 840
QY 835 LVHRDLAARNVLVSPHVKITDQGLARLLIDDETHADGCVPIKMMALLESILARRTT 894
DB 841 LVHRDLAARNVLVSPHVKITDQGLARLLIDDETHADGCVPIKMMALLESILARRTT 900
QY 895 HQSDVMSYGVTVLMTFGAKPYDGIIPAREIPDLLEKGERLPQCTIDVYMIWIKCM 954
DB 901 HQSDVMSYGVTVLMTFGAKPYDGIIPAREIPDLLEKGERLPQCTIDVYMIWIKCM 960
QY 955 IDSECRPRFELVSESRMARDPQRFVIVNEDLGPASPLDSTFFYSLLDEDDMGDLVDA 1014
DB 961 IDSECRPRFELVSESRMARDPQRFVIVNEDLGPASPLDSTFFYSLLDEDDMGDLVDA 1020
QY 1015 EBYLVPOQGFPCPDAGAGAMVHRHRSSTRSGGDLTLGLEPEEERAPSPPLASBGC 1074
DB 1201 EBYLVPOQGFPCPDAGAGAMVHRHRSSTRSGGDLTLGLEPEEERAPSPPLASBGC 108
QY 1075 AGSDVFPDGLGMAKALQSLPHDPSPLQRYSEDPVPLPSTDGVAPLRTSPQPEVY 113
DB 1081 AGSDVFPDGLGMAKATKQPSISPRDISPLQRYSEDPVPLPSTDGVAPLRTSPQPEVY 114
QY 1135 NQPDVPRQPPSPREGPFLARPAGATLIERAKTISPKNGVKQVAFAGAVENPEVLTQ 119
DB 1141 NQPDVPRQPPSPREGPFLARPAGATLIERAKTISPKNGVKQVAFAGAVENPEVLTQ 120
QY 1195 GGAAPQPPPPAPSPAFDNLVYWDQPPERGAAPSTFFKGTFAENPEVYGLDVPV 1249
DB 1201 GGSASQPH-PPALCPAFDNLVYWDQPPERGAAPSTFFKGTFAENPEVYGLDVPV 1254

```

RESULT 3
ERR2_RAT
ID ERR2_RAT STANDARD; PRT; 1257 AA.
AC P06494;
DT 01-JAN-1998 (Rel. 06, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor Protein-tyrosine kinase etdb-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor receptor-related protein).
CN ERBB2 OR NEU.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroblastoma;
RX MEDLINE=66118662; PubMed=3945311;
RA Bargmann C.I., Hung M.-C., Weinberg R.A.;
RT "The neu oncogene encodes an epidermal growth factor receptor-related protein."
RL Nature 319:226-230(1996).
RN [2]
RP SEQUENCE OF 852-905 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially expressed in the vertebrate nervous system."
RL Neuron 6:691-704(1991).
RN [3]
RP STRUCTURE BY NMR OF 650-668.
RX MEDLINE=92155181; PubMed=1346763;
RA Gallik W.J., Bottomley A.C., Lofis F.J., Doak D.G., Mulvey D., Newman R., Cumpston M.J., Sternberg M.J.E., Campbell I.D.;
RT "Three dimensional structure of the transmembrane region of the proto-oncogenic and oncogenic forms of the neu protein."
RL EMO J. 11:43-48(1992).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX, ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-ALPHA AND AMHRREGULIN.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS. THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation. CC The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/>)

OY 1133 YVNDPVPPEPPEEGPLPAARPGATLERAKTLSPGKGVKVDVFAFGAVENPEYLT 1192
 DB 1141 YVNOSEVOPPLTPEGLPEPVRAGATLTPKTLSPGKGVKVDVFAFGAVENPEYLV 1200
 OY 1193 FQGAAPQPHPPFPAFPAFDNLYYWDQDPPERGAPSTFGKTPTAENPEYLGIDVPV 1249
 DB 1201 PREGTASPFPHPSPAFSAFDNLYYWDQNSSEQGPPEPNEGFTGTAAENPEYLGIDVPV 1257
 RESULT 4
 EGFR_HUMAN STANDARD; PRT; 1210 AA.
 ID EGFR_HUMAN
 AC P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; O00732;
 AC O00688; Q9B2S2; Q9H2C9; Q9H2C9; Q9H3C9;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
 DE protein-tyrosine kinase ErbB-1).
 GN EGFR OR ERBB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CX NCB1_TaxId=9606;
 RN (1)
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=84219729; PubMed=6328312;
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
 RA Lee J., Varden Y., Libermann T.A., Schlessinger J., Downward J.,
 RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant
 RT expression of the amplified gene in A431 epidermoid carcinoma cells";
 RL Nature 309:418-425(1984).
 RN (2)
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Placenta;
 RX MEDLINE=95382957; PubMed=7654368;
 RA Ilekis J.V., Stark B.C., Scoccia B.;
 RT "Possible role of variant RNA transcripts in the regulation of
 RT epidermal growth factor receptor expression in human placenta";
 RL Mol. Reprod. Dev. 41:149-156(1995).
 RN (3)
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Placenta;
 RX MEDLINE=97078686; PubMed=918811;
 RA Reiter J.U., Maible N.J.;
 RT "A 1.8 kb alternative transcript from the human epidermal growth
 RT factor receptor gene encodes a truncated form of the receptor";
 RL Nucleic Acids Res. 24:4050-4056(1996).
 RN (4)
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Placenta;
 RX MEDLINE=97256547; PubMed=9103388;
 RA Ilekis J.V., Garcia J., Niederberger C., Scoccia B.;
 RT "Expression of a truncated epidermal growth factor receptor-like
 RT protein (TEGFR) in ovarian cancer";
 RL Gynecol. Oncol. 65:36-41(1997).
 RN (5)
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
 RC TISSUE=Placenta;
 RX MEDLINE=21100872; PubMed=11161733;
 RA Reiter J.U., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
 RA Schehl Sinclair C., Pearshall R.S., Green P.J., Yee D., Lampland A.L.,
 RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,
 RA Maible N.J.;
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative EGFR transcripts encoding truncated receptor
 RT isoforms";
 RL Genomics 71:1-20(2001).
 RN (6)
 RP SEQUENCE OF 575-687 FROM N.A.
 RA Reiter J.U., Threadgill D.W., Danielson A.J., Schehl C.M.,
 RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
 RA Maible N.J.;
 RT "Human and mouse alternative EGFR transcripts encoding only the
 RT extracellular domain of the receptor";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN (7)
 RP SEQUENCE OF 713-924 FROM N.A.
 RX MEDLINE=84196372; PubMed=6326261;
 RA Lin C.R., Chen W.S., Krutiger M., Stolarczyk L.S., Weber M.,
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
 RT "Expression cloning of human EGF receptor complementary DNA: gene
 RT amplification and three related messenger RNA products in A431
 RT cells";
 RL Science 224:843-848(1984).
 RN (8)
 RP SEQUENCE OF 150-962 FROM N.A.
 RX MEDLINE=84245835; PubMed=6305667;
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
 RA Roe B.A., Merlino G.T., Pastan I.;
 RT "Human epidermal growth factor receptor cDNA is homologous to a
 RT variety of RNAs overproduced in A431 carcinoma cells";
 RL Nature 309:806-810(1984).
 RN (9)
 RP SEQUENCE OF 1028-1210 FROM N.A.
 RX MEDLINE=85046483; PubMed=6093780;
 RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
 RA O'Kallie B.W.;
 RT "Isolation of an evolutionarily conserved epidermal growth factor
 RT receptor cDNA from human A431 carcinoma cells";
 RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
 RN (10)
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=88217333; PubMed=3329716;
 RA Haley J.D., Whittle N., Bennett P., Kitchington D., Ullrich A.,
 RA Waterfield M.D.;
 RT "The human EGF receptor gene: structure of the 110 kb locus and
 RT identification of sequences regulating its transcription";
 RL Oncogene Res. 1:375-396(1987).
 RN (11)
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=91107677; PubMed=1988448;
 RA Haley J.D., Waterfield M.D.;
 RT "Contributory effects of de novo transcription and premature
 RT transcript termination in the regulation of human epidermal growth
 RT factor receptor proto-oncogene RNA synthesis";
 RL J. Biol. Chem. 266:1746-1753(1991).
 RN (12)
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=85270438; PubMed=2991899;
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
 RT "Characterization and sequence of the promoter region of the human
 RT epidermal growth factor receptor gene";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
 RN (13)
 RP SEQUENCE OF 540.
 RA Kohda D.;
 RT Submitted (SEP-1997) to the SWISS-PROT data bank.
 RN (14)
 RP RECEPTOR ACTIVITY.
 RX MEDLINE=84191554; PubMed=6325948;
 RA Mroczkowski B., Mosig G., Cohen S.;
 RT "ATP-stimulated interaction between epidermal growth factor receptor
 RT and supercoiled DNA";
 RL Nature 309:270-273(1984).
 RN (15)
 RP PHOSPHORYLATION.
 RX MEDLINE=89278137; PubMed=2543678;
 RA Margolis B.L., Lax I., Kris R., Dombajagiam M., Honnegger A.M.,
 RA Hawk R., Givol D., Ullrich A., Schlessinger J.;
 RT "All autophosphorylation sites of epidermal growth factor (EGF)
 RT receptor and HER2/neu are located in their carboxyl-terminal tails.
 RT Identification of a novel site in EGF receptor";
 RL J. Biol. Chem. 264:10667-10671(1989).
 RN (16)

RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
 RP ASN-528.
 RX MEDLINE=96398132; PubMed=8962717;
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
 RT "Analysis of the glycosylation patterns of the extracellular domain of
 RT the epidermal growth factor receptor expressed in Chinese hamster
 RT ovary fibroblasts.";
 RL Growth factors 13:121-132(1996).
 (17)
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
 RP ASN-603.
 RX MEDLINE=20198209; PubMed=10731668;
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
 RT "Characterization of the N-oligosaccharides attached to the atypical
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor
 RT receptor.";
 RL J. Biochem. 127:65-72(2000).
 (18)
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE=98225196; PubMed=9556602;
 RA Abe Y., Odaoka M., Inagaki F., Lax I., Schlessinger J., Kohda D.;
 RT "Disulfide bond structure of human epidermal growth factor receptor";
 RL J. Biol. Chem. 273:11150-11157(1998).
 (19)
 RP REVIEW.
 RX MEDLINE=87297456; PubMed=3039909;
 RA Carpenter G.;
 RT "Receptors for epidermal growth factor and other polypeptide
 RT mitogens.";
 RL Annu. Rev. Biochem. 56:881-914(1987).
 (20)
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
 CC EGF-like growth factor, gp30 and vaccinia virus growth factor. Is
 CC involved in the control of cell growth and differentiation.
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
 CC secreted.
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p110 (shown here), 2/p60/
 CC truncated isoform/TBGR, 3/p110 and 4; are produced by
 CC alternative splicing.
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
 CC expressed in ovarian cancers.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X00588; CAA25240.1; -;
 DR EMBL; U95089; AAB53063.1; -;
 DR EMBL; U48722; AAC50802.1; -;
 DR EMBL; U48723; AAC50804.1; -;
 DR EMBL; U48724; AAC50796.1; -;
 DR EMBL; U48725; AAC50797.1; -;
 DR EMBL; U48726; AAC50798.1; -;
 CC -----
 Query Match 46.3%; Score 3133; DB 1; Length 1210;
 Best Local Similarity 49.4%; Pred. No. 1,3e-159; Indels 114; Gaps 23;
 Matches 626; Conservative 178; Mismatches 348;
 QY 11 LLLALLPPGAA-STGYC-----TGYIRANSKFTIGITELLYGQGVQVQGNLELYLPT 61
 Db 14 LLAALCPASRALERKVKVCGGTNSKLTQLGTGFEDHFLST-QRMENNCEVVLGNLEIIVYGR 72

QY 62 NASLSFLQDIOEVGVYVLIANQVQVPLQSLRIYRGTLQFEDNYALVLDNGPLNNTT 12
 Db 73 NYDLSFLKIQVAGVYVLIANTVERIPLLEMLQIRGMYYENSVALLVSNYD----- 12
 QY 122 PYTGAPEGGLRELQSLSTLEILKGGVLIQRNPOLCYQDTILMKDI FHKNNLATLIDTN 18
 Db 127 ----ANKTGLKELPMRNQOELIHGAVRFSNNPALCNVESIQMRDIVSSDFLSNMSDFQN 18
 QY 182 RSRACHPSGPMCKSRGCMGSESEDOGLSTRVACGGA-RKSGPLPDCHEGCAAGTG 24
 Db 183 HLGSCQKCDPSLCPNGSCWGAEBENQKLTIKICAQQSGRCRGSPSDCCNCAAGTG 24
 QY 241 PGRSDCLACLHFNHSGICELCPALVTYNTDTFESMPEDEGRYFGASCYACPNYLAST 30
 Db 243 PRESQCLVCRKRDATKDCQCPFLMLYNPTTYQMDVNPBEKGFSGATCYKKCRNRYVT 30
 QY 301 DVGSLTVLCPHLNDEVTLEDGTORCEKSKRCARVCYGLGMEHLREVAVTSANIQEPAG 36
 Db 303 DHGSCVRAAGADSYEM-EDGVKCKKCEGRCRCVNGDIGEFKDSLSTINATNIKHFKN 36
 QY 361 CKKIGSLAFLPESFQDPAANTAPLQPEOLQVETLEITGYVYISAMPDPLDLSVFO 42
 Db 362 CTSISGDHLIIPVARGDSFTHTPPLDQSLDITKYKEILGFLIQAWBENRTDLHAFE 42
 QY 421 NLQVIRGSLHNGAVSLTLQGLISWGLRSLRSLGSLALIHNNTHLCFVHTVPMQDLF 48
 Db 422 NLEIRGRKQHGQSLAVSLVSLNTISGLRELKESIDGDIISGNKOLCVANTINMKLLE 48
 QY 481 RNPQALHTANRPDEBEVSGIACRQLCARGHWGSGPQCVCNCSQFLNGQCEVEGCRV 54
 Db 482 GTSGGKTKIINRGENSKACGVCHALCSPEGWGEPFDVSCRNVSGRECVDCNLT 54
 QY 541 LQGLPREYVNAHRLCPHPECOPQNGSVTFGFPADQCVACAHYKDPFCVARGSPGVKP 60
 Db 542 LEGREPREVENSEICQHPBELPAMNITCGRPDNDICQAHYIDPHEVKTCPASVMG 60
 QY 601 DLSTYPIKFPDEBACQCPPTINCTHSCVDLDKCPAEPASPLTSIVAVYG---ILL 65
 Db 602 ENNTL-VKRYADAHVCHLCPHNTCYGTGSGLEGCPFNPKIP--SIATGMVALLLL 65
 QY 658 VVLGVNVLGILIKRROOKIRKYTRRLLOELVELETPSGAMPNQAMNLIKETELRY 71
 Db 659 VVALDIG---LPMRRHIVKRTIRRLQERBELVEPLTPSGAENQALLILETEFKI 71
 QY 718 KYLSGAGFVYKGIWIPQGENYKIPVAILVLRBNTPSKANKEILDEAYVMAGVSPYVS 77
 Db 716 KYLSGAGFVYKGIWIPQGENYKIPVAILVLRBNTPSKANKEILDEAYVMAGVSPYVS 77
 QY 778 RLIGICLSTVQLVLTQLMPYGCLLDHYREKRGSLGSDLLNMCQIAKMSYLEDVRLYH 83
 Db 776 RLIGICLSTVQLVLTQLMPYGCLLDHYREKRGSLGSDLLNMCQIAKMSYLEDVRLYH 83
 QY 838 RDLAARNVLYKSPNHYKITDFGLARLIDIEYEHADGGKVPYIKMMALLESILRRRTTHS 89
 Db 836 RDLAARNVLYKSPNHYKITDFGLARLIDIEYEHADGGKVPYIKMMALLESILRRRTTHS 89
 QY 898 DVNSGYVTWELMTFFGAKPYDGIIPAREIPLLEKGERLPPPICTIDVYIMVYKCMYDS 95
 Db 896 DVNSGYVTWELMTFFGAKPYDGIIPAREIPLLEKGERLPPPICTIDVYIMVYKCMYDS 95
 QY 958 ECRPREFELVSEFSRMRDPPORFVAVIO-NDLCPAPLSTPFRSLLEDODMDLYDAE 101
 Db 956 DSRKPRELIIERSKMARDPQRTLVYIQGDRKMLPSPPTSNFRALMDEEDMDVDVDAOE 101
 QY 1017 YLVQGGFPCFPDPAAGAGVHHRRHSSSTRGGGDUJLLEPSESEARSPPLAPSEGAG 107
 Db 1016 YLVQGGFPCFPDPAAGAGVHHRRHSSSTRGGGDUJLLEPSESEARSPPLAPSEGAG 107
 QY 1077 SDVEFDLIGMAAGKIGISLPTHPSPSLQRYSEDPYPLPSRT--DGVAFLPCLSGQPEVY 113
 Db 1042 SN-NSIVACIDRNGLOSCLIKDSFLQRYSSDPFGALTEDSIDDTFL-----PVPEYI 109


```

FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY) .
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY) .
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE)
FT CARBOHYD 128 128 N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT CARBOHYD 175 175 N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT CARBOHYD 196 196 N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT CARBOHYD 352 352 N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT CARBOHYD 413 413 N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT CARBOHYD 444 444 N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT CARBOHYD 528 528 N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT CARBOHYD 568 568 N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT CARBOHYD 603 603 N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT CARBOHYD 623 623 N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT CONFLICT 19 19 C -> S (IN REF. 2) .
FT CONFLICT 539 539 C -> W (IN REF. 5) .
FT CONFLICT 991 991 L -> F (IN REF. 4) .
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6) .
SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46DF2D2F5 CRC64;

```

Query Match 46.0%; Score 3115; DB 1; Length 1210;
 Best Local Similarity 49.5%; Pred. No. 1,2e-158;
 Matches 630; Conservative 170; Mismatches 355; Indels 118; Gaps 25;

```

QY 11 LLLALLPFGAA--STOYC-----TQYIKANSKFIGITELLYQGGQVQVQGNELTYLFT 61
DB 14 LITLALCAAGALAEKRCVCGCTSNRLTQLGTFEDHFLSL--GRMYNCEVVLGNLEITYVGR 72
QY 62 NASLSFQDIDQEVQGYVLIANOVQVPLQRLIRVSTQLEFEDNALAVLDNDPLANTT 121
DB 73 NYDSLFLKTIQEVAGYVLIANTVER:PLENLQIRGNALYENTYVLLALISN----- 124
QY 122 PVTGASPGGLRELQRELTLLKGGVLIQENPQLCVQDITLMDI----FKKNQALATL 177
DB 125 --YGTNRTGRELRLQELTIGAVFSNNPILCNNDITLQWMDIVQVFMSSMDL-- 180
QY 178 IDNRSBACHPCSPMCKGSRGCSSESDCSLTRTVQAGCA--RCKGPLPTDCCHQCA 236
DB 181 --QSHPSSCPKCDSPSCNGSCWGGGEEHCQKLTIKIICAQCSHRCGRSPDCHNQCA 238
QY 237 GCTGPKASDCLACHFNHSGICELHCPALVTYTTDFESNPNDEGRYTTGASCVTAQPN 296
DB 239 GCTGPRESDCLVCCQFQDEATCKDTCPLMLYNPTTYQMDVNEBGKYSFGATCVKCPFN 298
QY 297 YLSTDVGSCTLVCPILHNOQVTAEDGTQRCSEKSPCARVQVCGHMEHREVAVYASNIQ 356
DB 299 YVVTDHGSCYRACGPDYEV--BEDGIRKCKCQCPKRVNGGIGEFKDTLSINATNIK 357
QY 357 EFAGCKKIFGSLAFPSFGDPASNTAPLOPQLOVFEETLEITGYLYISAMPDLPDL 416
DB 358 HFKYCTAISGDLHLTPVAFKGDSEFTTRPPLDPRELEILTKVKEITGFLLIQAMPDMDTL 417
QY 417 SVFONLOVIRGRILHNQAVSLTLOGISWLGIRSLRELGSGLALHNHTHLCFNYTPW 476
DB 418 HAFENLEIIRGRTKQHQGFLLAVVGLNTLSLGRSLKEISDGVVITSGRNLCYANTINW 477
QY 477 DQLEFNHQAHLHTANRPEDECVGBGLACHQLCARHGCGPFTQCVNCSQFLRQCEVE 536
DB 478 KKLFGTQKTKIMNNAEKDCAVNHVCMPLCSSGCGMPEPRDVCQVNSRRECEVE 537
QY 537 ECRVLOGLPREYVNAHRCLEPCHCEQPOKNSYVCFEPREDQCYACHYKDPFCVACRPS 596
DB 538 KCNILBGEPRFEVNSECIQCHPEC:PPQAMNITCTGRGDNCTCQACHYIDGPHCKTQPA 597
QY 597 GVPKDSYMPKPFDEGACQPCPINCTHSCVDLDKGCSPAEOBRASPLTTS:VSAVGL 656
DB 598 GIMGENTTL--VMYYAADANNVCH:CHANCTYGCAGPQLOGCEVWPSPKPSIATATVIGGL 656
QY 657 LVVVLGVVFGI--LIKRCQKIRKTYTRRLLOETELVEPLTPSGAMPNQAKMLKTELAR 715
DB 657 LFTIV--VALIGLFMRRRHIVKRTLRRLLOERELVEPLTPSGEAPNQAHRIILKTEFEK 715
QY 716 KVKVLGSGAGTGYKGIWIPDGENVXI:PAIKVLRNENTSPKANKELDEAVYMAQVGSFY 775

```

```

DB 716 KIKVLGSAFGTGYKGLMIPEGEVXKI:PAIKELREATSPKANKELDEAVYMAQVSDNPH 775
QY 776 VSRLLGICITSTVQLVQLMPYGLDLHRENRQGLQSLWMCQIAGMSYLEDVRL 835
DB 776 VCRLLGICITSTVQLVQLMPYGLDLHREKDNIGSOTLWMCQIAGMNYLEDRL 835
QY 836 VHRLLARNVLYKSPNHVKTITDFGLARLDIDREYVADGAKYPIKMALESILRRPFT 895
DB 836 VHRLLARNVLYKTPQVHKITDGLAKLGAEREKTHAEGKXPIKMALESILHITHT 895
QY 896 QSDVMSGVTVYWEIMTFGAKPYDGI:PAEELPDLEKGERLPQPICTIDYVYIMVCKMI 955
DB 896 QSDVMSGVTVYWEIMTFGSKPYDGI:PAEELPDLEKGERLPQPICTIDYVYIMVCKMI 955
QY 956 DSECRPRELIVSEFSMARDDPORFVVIC--NEGLGASPLDSTFFYSLLDDMDGLVVA 101
DB 956 DADRREKREILIEFSQMAEDPORVLTIGQDERMHLPSPTDSNFRALMDDEBDEVDVA 101
QY 1015 EBYLVPOQGFPCDPAPAGAGVYHRRSSSTRSGGDLTLGLEPSEEEPARSPAPASEG 107
DB 1016 DEYLIPQGGFF-----NSPST-----SRTPLLSIS 104
QY 1075 AGSDVPFGDLMGAKGLQSLPTHDPSPLQVSEDTVPILPSET--DGVAAPLITCSPOPE 113
DB 1042 ATSN---NSTVACINENGSCRYKEDAFLOVRSDEPTGAVTEDNIDARL-----FVPE 109
QY 1133 YVNOVDVROPSPRESGLPAPAPACATLEBAKTLPGKNGVYKDVAFAGAVENPEYL- 119
DB 1092 YVNO-SVPKRPAQSVQVYVHNPPLP-----APGQDLHQN--PSNNAVNEPYLN 114
QY 1192 TPQGAAPQHPPPAPAFADNLVYWDQ-----DP-----PERGAPSTFKGTP 123
DB 1141 TAQ-----PTCSSGFNSPALMIQKSHQMSLDNDPDYQDDFFPKETKPGIRKQ-P 119
QY 1236 TAENPEYLG:VLP 1248
DB 1191 TABNAETLRVAPP 1203

```

RESULT 6
 EBB4_HUMAN
 ID_EBB4_HUMAN STANDARD; PRT; 1308 AA.
 AC Q15303;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor Protein-tyrosine kinase ebb4-4 precursor (EC 2.7.1.112) (p180erb4) (Tyrosine kinase-type cell surface receptor HER4).
 GN EBB4 OR HER4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM JM-A).
 RC TISSUE=Breast Carcinoma;
 RX MEDLINE=93189574; PubMed=8383326;
 RA Plowman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W., Foy L., Neuhauer M.G., Shoyab M.;
 RT "Ligand-specific activation of HER4/p180erb4, a fourth member of the epidermal growth factor receptor family";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750 (1993).
 RP SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
 RC TISSUE=Fetal brain;
 RX MEDLINE=9476287; PubMed=9334263;
 RA Elenius K., Corfas G., Paul S., Choi C.T., Rio C., Plowman G.D., Klagesbrun M.;
 RT "A novel junctional membrane domain isoform of HER4/erbB4. Isoform-specific tissue distribution and differential processing in response to phorbol ester";
 RL J. Biol. Chem. 272:26761-26768 (1997).

CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BECAMELULIN AND
CC NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: JM-A (SHOWN HERE) AND JM-B;
CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
CC KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID CEREBELLUM,
CC PITUITARY SPLEEN, TESTIS AND PANCREAS. LOWER LEVELS IN THYMUS,
CC LUNG, SALIVARY GLAND, AND PANCREAS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L07868; AAB59446.1; -
DR HSSP: P11362; 1FGK.
DR Genew: HGNC:3432; ERBB4.
DR MIM: 600543; -
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_pkinase.
DR InterPro: IPR004019; YLP_motif.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR Pfam: PF02757; YLP; 2.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SMO0261; FU; 4.
DR SMART: SMO0219; TYRC; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane: Glycoprotein; Multigene family; Receptor; Signal;
KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Alternative splicing.
FT SIGNAL 1 25
FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 652 675 POTENTIAL.
FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 486 633 CYS-RICH.
FT DOMAIN 718 985 PROTEIN KINASE.
FT NP_BIND 724 732 ATP (BY SIMILARITY).
FT BINDING 751 751 ATP (BY SIMILARITY).
FT ACT_SITE 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.

FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 617 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC...) (POTENTIAL).
FT VARSPPLIC 626 648 NGPISHCIYPPVGHSTLPQHA -> IGSSIEDICGLMD
FT (IN ISOFORM JM-B).
SQ SEQUENCE 1308 AA; 146807 MW; 5FAEB0985D86761 CRC64;
Query Match 43.8%; Score 2968.5; DB 1; Length 1308;
Best Local Similarity 45.1%; Pred. No. 8,4e-151;
Matches 609; Conservative 186; Mismatches 369; Indels 185; Gaps 30
9 WGLLALLPQGA-----STQCYGYIANKSKFGITEL-----LYGCGVVGQNL 55
8 WWSVSLVAAGTVPDSQSYCA---GTENKLSLSLDAQYRALRYRNCVWGNLE 64
56 LTYLPTNASLFLDIOEVQGVYLAHNOVROVPLRLIRVGTOLFEDNYALAVLNDG 115
65 ITSIEHNHDLFLRSVREVTGVYVVALNQRYPVLEWLRIRRGKLYEDRYALAFNYR 124
116 PLNNTPTPTGASPGGLBELQRLSTTELLKGVLIQRPOLCYODTILMKDIFPKNOAL 175
125 KDGNF-----GLQELGLKNTLETLLNGVYVDQKFLCYADTIIHWQDIVNPMPSNL 175
176 TLIDNRSRACHPCSPMCKGRGSESSDCQSLTRIVCAAGC-ARCKGFLPTDCHEQC 234
176 TLVSINGSSGGGRCHKSGTQ-RCMGPLENNCQITLIRIVCAEQCDGRGVGVSDCHREC 234
235 AAGCTGPRHSGCLALHPNHSGLTEIHCPLALVTYNDTFESMNPBGRVYFGASCYTACP 234
235 AAGCGSPDTCFCACMFNDSGACVTCQPTFYVNPPTFLPHFNFAKYVGAFCVKKCP 294
295 YNYSSTDVGSCTLVCPPLHNOVTAEDTQGCCKSKRCVAVCYGLGHEHREVAVTSAN 354
295 HNFV-VBSSSVRACPSRKNV-EENGIWKCKRPTDLCPRACGIGISGLMSQYDSSN 352
355 IOEPAGCKKIGSLAFIPESFDGDPASNTAPLOEQIOVEFTELEITGYLYISAMPDSEP 414
353 IDKFNCTKIKINGNLIFFVTGIGHDPYVAIAIDBEKINVRTRYREITGFANISMPNMT 412
415 DLVFNQNLQVIRGRILHNGAVSLTLOGLGSMIGRLRELSGLLIHNHTLCKVHTV 474
413 DPSVFSNLVTIGGVYLSGLSLKLGQGLTSLQFOSLKEISAGNIYITDNSVLCTYHTI 472
475 FMDQLFRNPQALJHTANRPEDECVGEGLACHQICARGHCWGPPTQCVNCSQFLRGQEC 534
473 NMTLFTSTINQRIYIRBNRKAENCTAEGVCMNLCSDDGCMWGPQDCSCRFSGRIC 532
535 VEEGRVLDGLPREVFNABHCLPCHPECP-QNSVYVCFGEALDQCACAHYKDPPTCVAR 593


```

Db      533 IESCNLYDGEPEFENGSI CVECDPOCEMEDGILITGHPGPDNCTKSHKFNPCNVEK 592
Qy      594 CFSGVKPDLSYMPIMKFPDEEGACOPCEINCTHSCVDLDDKGC-----FAQRA 642
Db      593 CPDGLGANSF--IKRYADPRRECHPHCPNCTGCGNGSTSHDCIYIPWTGHTIPQAR- 649
Qy      643 SPLTSTVSANV-GILLVVLVGVFGILKKRQOKTRKTYMRLLQETELVEPLTPSGAMP 701
Db      650 TPL--IAAGVIGGLILVIVGLTFVAVVVRKSIK-KKRALRFL-ETELVEPLTPSGTAP 705
Qy      702 NOAOMRIKETELRYKVKYLGSGAFGVKGIWIPDGEVAKIPIVAKYLRNTSKAKKEI 761
Db      706 NQAQRLIKETELKVKVYLGSGAFGVKGIWIPDGEVAKIPIVAKYLRNTSKAKKEI 765
Qy      762 LDEAYVAGVSGPVYSLIGLICITVOLTVMPEYGLLDHVENRGRLGSDLLNWCN 821
Db      766 MDEALIMASMDPHLVRLIGVCLSTPTQLVTLQPMHGCILEYVHEHKDNIGSQLLINWCV 825
Qy      822 QIAKGMVSYLEDVRLVHRDLAARNVLYKSPNHVKITDTEGLAFLDIDSTEYHADGKVPK 881
Db      826 QIAKGMVSYLEDVRLVHRDLAARNVLYKSPNHVKITDTEGLAFLDIDSTEYHADGKVPK 885
Qy      882 WMALESILRRRTHSGDYSYGVYVWELMTFGANPYDGIPIAREIPDLLEKGERLPQPIC 941
Db      886 WMALECIHYRKEFTHQSDVMSYGVITWELMTFGKPYDGIPIAREIPDLLEKGERLPQPIC 945
Qy      942 TIDVYMWKCMIDSEGRPRFRELVSFSMADPQAFVITQNE-DLGPASPLDSTFYR 1000
Db      946 TIDVYMWKCMIDSEGRPRFRELVSFSMADPQAFVITQNE-DLGPASPLDSTFYR 1005
Qy      1001 SLLEDMDGDLVDAAEYLVPOGFGCPDPAAGAGAMVHRRSSSTRSGGDLTLGLEPS 1060
Db      1006 NLLDEEDLEDMDAAEYLV-PQAFNIIPP-----ITSRAKILSNNS-----EIGHSP 1053
Qy      1061 BEAPRS-----PLAP-SEGAGSVDFGDLGMAKAG 1091
Db      1054 PAYTWSGNQFYVRDGFPAEQGVSVPYRAFTSTIPAPVAGQATAIIFDSCNGTLRK 1113
Qy      1092 LQSLPTDPSPLQRYSEDPPTVPLPS-----ETDGVAPITCSPOPEYVNOGDVAPQPP 1144
Db      1114 PVAPHVQEDSSTQRYADPTVFAPERSRGRGLDEGIMTWRDQPKQOEYINPYE----- 1167
Qy      1145 SPREGPPLPAPPAQATLERAKTISPGKNGVYKDVAFAGAVENPEYLLTPOGGAAPQPHP 1204
Db      1168 --ENPFVSR--KNGDLQ-----ALDNEVYNAAG-----PP 1196
Qy      1205 PA-----FSPAFDNLVYVDQDPPERGA--PP 1228
Db      1197 KAEDRYNEPLVYLTNPANTLGAELVKNLISMEKAKKAFDNDVWNHSLPPRSTLQHP 1256
Qy      1229 STFQGTPT-----AENPEYL 1243
Db      1257 DYLOEYSTKYFYKONGRIPIVAPNPEYL 1285

```

```

RA      Marchionni M.A., Kelly R.A.;
RT      "Neuregulins promote survival and growth of cardiac myocytes.
RT      Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT      ventricular myocytes.";
RL      J. Biol. Chem. 273:10261-10269(1998).
RN      (2)
RP      SEQUENCE OF 848-901 FROM N.A.
RC      TISSUE=sciatic nerve;
RX      MEDLINE=91222560; PubMed=2025425;
RA      Lai C., Lemke G.;
RT      "An extended family of protein-tyrosine kinase genes differentially
RT      expressed in the vertebrate nervous system.";
RL      Neuron 6:691-704(1991).
RN      (3)
RP      SEQUENCE OF 1031-1198 FROM N.A.
RC      STRAIN=Sprague-Dawley; Tissue=spinal cord;
RX      MEDLINE=97184212; PubMed=9030624;
RA      Carroll S.L., Miller M.L., Fournier P.W., Kim S.S., Corbett J.A.;
RT      Expression of neuregulins and their putative receptors, ErbB2 and
RT      ErbB3, is induced during Wallerian degeneration.";
RL      J. Neurosci. 17:1642-1659(1997).
CC      - FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC      2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC      NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC      - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      - SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC      RECEPTORS (POTENTIAL).
CC      - SUBCELLULAR LOCATION: Type I membrane protein.
CC      - TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC      NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC      OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC      RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC      HEART.
CC      - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC      RESIDUES (BY SIMILARITY).
CC      - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      This SWISS-PROT entry is copyright. It is produced through a collaborative
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: AF041838; AAC08899.1;
CC      EMBL: U52531; AAC53051.1;
CC      HSSP: P11362; IFCK.
DR      InterPro: IPR000494; EGFR_L domain.
DR      InterPro: IPR000719; Euk_pkinase.
DR      InterPro: IPR002174; Furin-like.
DR      InterPro: IPR001245; Tyr_pkinase.
DR      InterPro: IPR004019; VLP_motif.
DR      Pfam: PF00757; Furin-like; 1.
DR      Pfam: PF00068; Pkinase; 1.
DR      Pfam: PF01030; Recep_L_domain; 2.
DR      Pfam: PF02753; VLP; 2.
DR      PRINTS: PR00109; TYRKINASE.
DR      PRODOM: PD000001; Euk_pkinase; 1.
DR      SMART: SMO0261; FU; 4.
DR      SMART: SMO0219; TYRK; 1.
DR      PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR      PROSITE: PS0011; PROTEIN_KINASE_DOM; 1.
DR      PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW      Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW      Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT      SIGNAL 1 25
FT      CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT      DOMAIN 26 651
FT      TRANSMEM 652 675 POTENTIAL.
FT      DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).

```

```

FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 CYS-RICH.
FT DOMAIN 718 985 PROTEIN KINASE.
FT NP_BIND 724 732 ATP (BY SIMILARITY).
FT BINDING 751 751 ATP (BY SIMILARITY).
FT ACT_SITE 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOXYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOXYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOXYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOXYD 358 358 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOXYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOXYD 473 473 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOXYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOXYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOXYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOXYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOXYD 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 PYR -> SYR (IN REF. 3).
SQ SEQUENCE 1308 AA; 146957 MW; D948B096A08B41 CRC64;

```

Query Match 43.5%; Score 2949; DB 1; Length 1308;
 Best Local Similarity 45.0%; Pred. No. 9, 2e-150;
 Matches 607; Conservative 193; Mismatches 378; Indels 172; Gaps 30;

```

QY 1 MEUA-ALCEWGLL--ALPPGASTOVCTQYIKANSKFIGITEL-----LYGCGQ 48
D 1 MKLATGLMWGSLLVAAFTVQPSASGVCA---GTENKLSISLDEQYRALRKTYENCE 57
QY 49 VVUGNLELYLPINNASLFLQDIQEVGYVLIHANQVQVLPQRLIRVGTQLEFNVAL 108
D 58 VVMGNLEIITSIBENRDLSEIRSVETGYVALNQFRYLPLENIRIRGKLYEDRYAL 117
QY 109 AVLDNGDPLNNTTPVYGASPGG:REIOLSLTEILKXGVLIOBNQOLCYOCTILMKPIFH 168
D 118 AIFLNYRKQGNF-----GLQELGKNIIEILANGVYVDQNFLECYADTIHMQDIVR 168
QY 169 KNNQALTLITIDNRSACPCSPMCKGSRFCWGESSEDCOSLTRVYACAGC-ARCKGPLPT 227
D 169 NPWPSNMTLVSTIGSGCGCRCHKSGCTG-RCWGPTENHCQTLIRTVYACQCGRCYGPYVS 227
QY 228 DCCHQCAAGCTGPKKSDCLACLFHNSGICELHPALVTYNTDFFESMPNPEGRYIFGA 287
D 228 DCCHRECAAGCGSPKDTDCFACGMNNDGACVTCQPCQTFVYNPTTFQLEHNFNKRYTGA 287
QY 288 SCVYACPNYVLTVDGCTLVCP:LANQEVTAEDGTQRCSEKSKPCARCYGIGMEHLREV 347
D 288 FCVKCKPBNFV-VDSSSCYRACPCSSKMEV-BENGIMCKMCKPCTDIOCPYACDGIIGLSLMSA 345

```

```

QY 348 RAYTSANIOEFACCKIFGSLAFIPSPFGDASNTAP:QPELOQVFTLEBITGYLIS 40
D 346 QYDSSNIDDFINCTKINGNLIFLVYTGHDPPYNAIDALDPEKLVNFRVREITGFLNQ 40
QY 408 AMPDSLPLDSVPONLOVIRGRILHNGAYSLTQGLGISWGLRSIRELSGALLHHNTH 46
D 406 TWPNMIDFVFENLVITIGRIVYSLSLILKQOSITSLQFOSLKEIAGNVIYIDNSN 46
QY 468 LCFVHTVPMQDLFRNPHQALLTANRPEDECVGEGIACHQLCARGHCMWPGPPTQVCNSQ 52
D 466 LCYHTINMTTLTSTYNORIVIRDNRAENCAEGVNCNHLCSNGCMWPGPDQCLSCR 52
QY 528 FLRGQECVEECRVLQGLPREVYNARHCLECHCECP-QNGSVTCCGPREDDQVACAHYD 56
D 526 FSGKICIESCNLYDEFFREFENGSI:CVCDSDCEMEDEGLTCHGPGPDNCTKCSHFED 58
QY 587 PFCVACRPSGVKPDLSYMPIMKFPDEGACQPCPINCTHSCVDDLDDKGC----- 63
D 586 GPNCEVKCPDVLQGANSF--IFKYADQDECHCHCHNCTQCGNCPHSHDCIYYPMTHST 64
QY 637 -PAEQASPLTISVSAV-:CILLVYV:GVVPEGILIKRQOKIRKTYTRRLLOETLVEPL 69
D 644 LPQPAR-TPL--IAAGVIGLFLVIMALTFAVYVRRSIR-KKRALRRL-ETELVPEPL 69
QY 695 TPSCAMPNQAQMRILKETELRKVYLGSGAFGTYYKGIWIPDGENYKI:PVAIKVRENTS 75
D 699 TPSCGTPNQAQRLKETELRKVYLGSGAFGTYYKGIWIPDGENYKI:PVAIKIINETTG 75
QY 755 PRANKETLDEAYMAGVSPYVSRLLIGTISTVOLVTQLMRYGCLLDHVRNRRGLSQ 81
D 759 PRANKVEFMEBALMASVDHPHLVRLIGVCLSPITQVLQVLMHGCLLEVYEHKONISQ 81
QY 815 DLNMCQIAKGSYEDVRLVHRDLAANVYKSPNHYKIDPGLAALDDDEFENAD 87
D 819 LILNMCVQIAKGMVIEERLVRHDLAANVYKSPNHYKIDPGLAALDEDEEVAD 87
QY 875 GGVKPIKMALESILRRPFTQSDVSYGVTVMEITGAKPYDGI:PAEIRPDLLEKER 93
D 879 GGVKPIKMALESILRRPFTQSDVSYGVTVMEITGAKPYDGI:PAEIRPDLLEKER 93
QY 935 LPQPICTIDVYMIWKKCMWIDSECRPREFRELVSFSMADDPQRFVYIQNED-LGPAASP 99
D 939 LPQPICTIDVYMIWKKCMWIDSDSRPFKEIAAEFSMARDPQRYVLIQDDRWKLPSP 99
QY 994 LOSTFYRLLEDMDMDGLVDABEYLVPOQGFPCDP-----A 10
D 999 NDSKFPQNLDEDELEDMDABEYLV-QANVIPPITYTSRTRIDNSSEIGHSPRAYT 10
QY 1031 PGNAGVYHHRHSSSTRSGGDLTGLBESSEEAFRSLAPSEGAGSDVPDGLMGMAK 10
D 1058 PMSGGVYVQDGGFAITQGG--MPVPYATITSTIPEAFAV--QGATVEMFDSCCNGLTR 11
QY 1091 GLQSLPETHDPSPLQRYSEDPYVFLPS-----ENDGVVAPLQTSPOREYVNOQDVNRPQ 11
D 1113 KPVPHVVOGDSSTQRYSDAPYFABERPRABLBEGMTWMDKFKQEVYANPAYE----- 11
QY 1144 PSBRPCLPAAFPAGATLERAKTSLSPGNQYVKQVFAFGAVENPEYLLTPQGAAPQHP 120
D 1168 -----ENPFVSR-----KNGDLO-----ALDNPEVHASASG-----P 11
QY 1204 PPA-----FSPAFLDLYWDDQPPRGA--P 12
D 1196 PPADEBYVMEPLYNTFTNALGNMEIYMNKSLISPEBAKKAFFNDPMWNSLPERSTLOH 124
QY 1228 PSTFGKPT-----AENPEYL 1243
D 1256 PDLQYESTKYFYKONGRIRPIVAENPEYL 1285

```

RESULT 8
 XMRK_XIIPMA
 ID XMRK_XIIPMA
 AC P13388;
 STANDARD; PRT; 1167 AA.

Query Match	39.8%	Score 2696.5	DB 1	Length 1167
Best Local Similarity	45.5%	Pred. No. 2.3e-136		
Matches 575	Conservative 163	Mismatches 388	Indels 139	Gaps 26
4 AALCRWGLIALLPKANST---OVC-----TOYIKANSKEFIGETELLYOGCQVYGNLE	55			
8 AALIQ--LLILVLSIRKCSCTDPDRKCYCGSTNQMTMLDNHLYKKK--MYGCVNVLIENTLE	64			
56 LTVLPTNASLSPFODIOEVGYVLLAHNQVQVPLQRLRIYRGQLFEDNYALAVDNGD	115			
65 ITTQENQDLSFQISIQEVGYVLLAHNEVSTPLVNLRLRGNLLEGNFTLLVMSNY	124			
116 PLNNTPTVYGASPGGLRELQRLSLTELKGVLLIQRPOLCYOPTILMKDIFPKNQAL	175			
125 K-NPSPF--DYQVGLKQLQSLNLELISGVKYSHPMLCNVETIMMVIDKTSNPTM	181			
176 TLIDTKRSARCPGSCMGKSGKSCWSESSEDQSLTRYVCAGCC-ARCKGPLPDDCCEOC	234			
182 NLTPHAFERQCKCHGCGVNGSCVAPRGHQKPTKLICAEQCRRCRGKPRIDCCNEHC	241			
235 AAGCTGPKRSDCIACILHFNHSGICELHCPALVYVNTPTFSMPREGRYFGASCTACG	294			
242 AGGCTPRATDCLACRDNDDGCTCDTPPKIYDIYSHQVNDENPKYTFGLACVYCEB	301			
295 YNLTSDVSSCTLVCPRLNHQEVTAEDGTQRCCKSKPCARVCYGLGMENHLEVRVATSAN	354			
302 SNVYVE--GACVRSASGMLLEV-DENGKRSCKPDCGVCPKCDGIGISLNTIAVNSTN	359			
355 IOFPACCKIFPSGLFLFESFGDPASNTARLOEIOVETLEBETIGYLYISAMPDISP	414			
360 IREFSNCTKINDITILNRSFEGDHHYKIGTMDENHLMNLTVEKLTGYLVIMMVENNT	419			
415 DLSVFONTLVIRGRILHNHGAYS--LTLQGLISWYGLSLRELISGLALIHNNHLCFYNT	473			
420 SLISVFONTLEIIIGRTTFSRGSFVYVQRLHQMGLRSLEVASGAVILKNTIQRLRYAT	479			
474 VPPDQLFRPHRALHTANPREDECSGEGGLACHOLCARGHCWGPRTQVNCNSQFLRGE	533			
480 IMRRIRFRESDSLEYDART-----ENDTCNNECEDCWPGRPTMCSCLHVDGGR	532			
534 CVEECKVYGLPREVYNARHCLPCHPECPQNSGVTFFGEADQCVACAHYKDRPCVAR	593			
533 CVASCNLLDGEEREAQVDRGVQGHQECIVQTSILTYGGRANCKSAHFQDGRPCR	592			
594 GSGAKRDT SWKATIKYFEDTSCAGGQVYVNTQNTLITWGLVNTVNTVNTVNTVNTVNT	593			

```

Db 593 CPHGILGSDPTL-IMKXADKMGQCQCHQCTGCGGPGCLSGCRGP-LYSHSLAVGLVS 650
Qy 654 GILLVVLVGVFGILLIKRQOKRKTKTMRLLQETLVLPPLPSGAMPQOAMRLIKTE 713
Db 651 GLLITVALLIVLLRRRRIRK-RKRTIRCLLOEKELVPLPPSOQAPQOATRLIKTE 709
Qy 714 LRKVKVLGSGAGFTVYKGIWIPGGEVAKI PVAIKVLRNTSPKANKELIDEYVWAGVGS 773
Db 710 FKQDRVLGSGAGFTVYKGLMNPGENIRIPVAIKVLRNTSPKANKELIDEYVWAGVGS 769
Qy 774 PYVSRLLGICLTSTVOLVQLMYGGCLLDVHNRGRGLSGQDLNMCQIAKMSYLEDV 833
Db 770 PIVCRLLGICLTSAVOLVQLMYGGCLLDVHNRGRGLSGQDLNMCQIAKMSYLEER 829
Qy 834 RLVAHRLAARNVIVKSPNPKIKIDFGLRLDIDETEVYADGKVPKIMMALESLTRRF 893
Db 830 HLVAHRLAARNVILKSPNPKIKIDFGLSKLLTADKEVYADGKVPKIMMALESLTRRF 889
Qy 894 THQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYIMVWCM 953
Db 890 THQSDVWSYGVTVWELMTFGSKPYDGI PAKETIASVLENGERLPQPEICTIEYVMIILKCM 949
Qy 954 MIDSECRPFRELVSEFSMARPDQSFVVIQONEDLGPAASPLDSTYRSLDDDDMGDLVD 1013
Db 950 MIDPSRPFRELVSEFSQMARDPSTYVIQ--NLPSLSRRLFSLSSDD--DVVD 1004
Qy 1014 ABEYLVPOQGFPCPDPAQAGGVHRRHRSSTRSGGDLTLGLPSESEAPRSLAFSE 1073
Db 1005 ADEYLLPYKRI-----NRQSS-----EKCIPPT 1027
Qy 1074 GAGSDVFDGDLGMAKAGIQLPHTDPSFLQRYSEDPV-PLPSTDDGVAPLTCSPQE 1132
Db 1028 GH-----PVRENSTLTJLNISDPQNALKXLDH-----E 1057
Qy 1133 YVNPDPVRPQ-----PSPRE-----GPLP-AAAPGATLERAKTUSPGKGVVXDF 1179
Db 1058 YVNPQSESTSRSLDLYNPNYEDLTGCMGPVLSQGEATNRSREYNTNONS---PL 1114
Qy 1180 AFGAVENPEYLTPOGGAAPQHPPPAPSPADNLVYWDQDPPERGAPPSFFKGTPTAEN 1239
Db 1115 VSSGSMDDPDY---QAG-----YQAAF-----LPQTGALTGMFLPAEN 1152
Qy 1240 PEYLG 1244
Db 1153 LEYLG 1157

```

```

RA pioman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M., expression of an additional epidermal growth
RT "molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909 (1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=placenta;
RX MEDLINE=9328282; PubMed=7685152;
RA Kachon W., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197 (1993).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULIN AND NTK.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaborati
CC between the Swiss Institute of Bioinformatics and the EMBL outstation i
CC the European Bioinformatics Institute. There are no restrictions on i
CC use by non-profit institutions as long as its content is in no w
CC modified and this statement is not removed. Usage by and for commerci
CC entities requires a license agreement (see http://www.1sb-sib.ch/announc
CC or send an email to license@1sb-sib.ch).
CC
DR EMBL; M29366; AAA35790.1; -
DR EMBL; M4309; AAA35979.1; -
DR EMBL; S61953; AAB26935.1; -
DR PIR; A36223; A36223.
DR HSSP; P11362; 1FGK.
DR Genew; HGNC:3431; ERBB3.
DR MIM; 190151; -
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU_3_Pkinase; 1.
DR SMART; SM00219; TykC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP_FALSE_NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR_FALSE_NEG.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM_1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Alternative splicing.
FT SIGNAL 1..19
FT CHAIN 20..1342
FT DOMAIN 20..643
FT TRANSMEM 644..664
FT DOMAIN 665..1342
FT DOMAIN 709..966
FT NP_BIND 715..723
FT BINDING 742..742
FT ACT_SITE 934..934
FT DISULFID 186..194
FT DISULFID 190..202

```

```

RESULT 9
ERBB3_HUMAN STANDARD; PRT; 1342 AA.
ID ERBB3_HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERBB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERBB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197 (1989).
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;

```

```

FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.
FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 286 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 552 565 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 141 183 EILSGVYIEKNDLCMDTIDMDIWDIADDAEIVKDNKR
SC -> GQPFVPSGLTPROADMDYLLDDPRRLTLTASASK
VPTVLAIV (IN SHORT ISOFORM).
FT VARSPLIC 184 1342 MISSING (IN SHORT ISOFORM).
FT CONFLICT 560 560 E -> G (IN REF. 2).
FT CONFLICT 1064 1064 E -> G (IN REF. 2).
SQ SEQUENCE 1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;

```

Query Match 35.8%; Score 2423; DB 1; Length 1342;
 Best Local Similarity 40.7%; Pred. No. 9, 7e-122;
 Matches 532; Conservative 190; Mismatches 446; Indels 140; Gaps 33;

```

QY 5 ALGRWGLLALLPFGAASVQYQYIKANSKFGITELLVQGGQVYQGNLELTYLPTNAS 64
DB 27 AVCP-GTLNGLSVTDENQYQYLY-----KLYERCEVVMGNLEIVLTGNMAD 73
QY 65 LSFLODIOEVGYVLIANQVQVLOQLRIYRGTOLEFENNVALALVDNDPLANTTPTV 124
DB 74 LSFLOMIREVTGYLVAMNESTPLPLNLRVVGSTGYDGGKFAIFVM-----LVNVT-- 125
QY 125 GASPDGLRELQRLSTLEILKGVLIQENPOLCYODTILMKDIFHKNOQLATLTIDTNR 184
DB 126 -NSHALRQLRLTQLEILSGVYIEKNDLCMDTIDMDIWDIADDAEIVKDNKR 181
QY 185 ACHFCSPMKCKSRWSSSSDQSLRTTVAGGC-ARCKGRLPTDCCHQCAAGCTGPRH 243
DB 182 SCPPEHEVCKG-RWGGSGSDCQTLTKTICAPQNGHCFEPFNQCCHCEGAGSGGSPD 240
QY 244 SDCLACLFHNSGICEHCPALVTYNTDFESWPNREGRTFGASCVTACPVNYLSTDVG 303
DB 241 TDCFACHFNDSGACVRCRCPPLVYNKLTQLEPNHTKQYGVGVASCFHNFV-VDOT 299
QY 304 SCTLVCPLNQVETAEDEGTQRCCKSPCARVYGLGMEHLEVRVAVTSANIQEFAGCK 363
DB 300 SCVACPPDKMEVD-KGLMKCEPCGGLCPKACEGTGSG--SRFQTVDSNIDGFVNCTK 356
QY 364 IFGSLAFLPESPDPPASNTAPLOPELOVFELEITGYLVISAMPDLSPLSLYFQNLQ 423
DB 357 ILGNLDFLITGLNDPNNKIPALDPEKLVFRVREITGYLVNIQSPPPMHFSVFSNLT 416
QY 424 VIRGRILHNGAYS-LTLQGLISWLGRLSIRELGSGLAIHNNTHLCFVHTVPMQDLFRN 482
DB 417 TIGRSLYNRGFSLLIKGNLNTVSLGFRSLKETISAGRITYISANRQICVHNSLWTKVLRG 476

```

```

QY 483 PHOALHTA-NRPEDECVGEGLACHOLCARGHCWGPRPTQCVNCSQFLRGQVEEGRVL 541
DB 477 PTERLIDIKHRRPRDCVABSKVCDPICSSGGCMGRPGQCLCRNNRSRGVCVTHCNFL 536
QY 542 QGLPREYVNAHCLPCHPECOFQNGSVTCTGPEADQCVACAHYKDPFCVACRPSGVKPD 601
DB 537 NGEPREFAHEACSCSHPCEQCPMEGTATCNGSGSDTCAQCAPHRDGPHCVSSCPHGLG- 595
QY 602 LSYPMKWPPEEGACQCPINCNSGVDDLDDBGCPRAEGA-----SPLTSIVAAVYGLL 657
DB 586 -AKGPIYKRPVQNECPCHENCTQCKGPELDQGLVLIGKHTLTAALVIAVG-L 652
QY 658 VVVLGVVFGILIKRQOKIR-KYTMRLLOETELVEFLTPSGAMPQACWRLIKETELRK 716
DB 653 VVIFMVLGGTFLYVRGRIRQNKAMRRYELRGSIIEFLDPS-EKANVLAIRFKETELRK 711
QY 717 VKVLGSAFGVYVYGVITIPQGENVKIYALKVLPENTSPRANKIIDEAVYMAVGSPVY 776
DB 712 LKVLGSGVFTVHGVMIPGESIKIPVCIKVIEDSKGRQSPQAVVDHMLAIGSLDHAH 771
QY 777 SRLIGICTSTVQLVLTQMPYGLLDHVENRRLGSGDLNMCQIAKMSYLEVRLV 836
DB 772 VRLIGLCRGSSLOLVTOYVLPGLSLDHYRQKRALGSQLLNNGVQIAKMYLIEHGVY 831
QY 837 HRDLAARNVLYKSPNHVKITDFGLARLLIDETRYADGSKVDIKMMALESILRRFTHQ 896
DB 832 HRNLAARNVLYKSPQVAVDFGVADLLPPDDQLLYSEAKTPIKMAALESIFHGKYTHQ 891
QY 897 SDVMSYGVYVWELMTFGAKPYDGPAREIPDLLEKBERLPORPITIDYVMVVKCMMD 956
DB 892 SDVMSYGVYVWELMTFGAEYPAGIYLAIEVDLLEKGRLAQPOICTIDYVMVVKCMMD 951
QY 957 SECPREFRELVSERSMARDPORFVYIQLNEDLPA---SPLDSTFYRSLLEDDMDGLVD 101
DB 952 ENIRPTKEILANETRRAPDPRYLIVIKRES-GPGLAAGEPHGLNKKLEBELLEPDL 101
QY 1014 AEYLTVQQGFPCDPAPRAGAGVNHHRNSSSTRSGGDLTLGLEP-SEEEAPRSLAPS 107
DB 1011 LDLDLEAED-----NLATTTGSAISLTVGTJLRPRGQSLSLSPS 105
QY 1073 EGASDVDFDDLGMAKGLQSLPTHD-PSPLQRYSEBDTPVLP-----SENDGYVA-- 112
DB 1052 SGY-MPMNQNLJESCOESAVSSSERCPRVSLH-----EPRCCLASSESGHVTGS 110
QY 1124 -----PLTCSPOPE-----YNNQPDVRDQPSPREGP-----LPRA 115
DB 1105 EAELOEKVSMCRSSRSRSPRPGSDAHSQRHSLTPTVPLSPGLSEEDVNGVYMPDT 116
QY 1155 RPAGATLERAKTILSP-GKNGVY-----KDVFAFGAVENPEVLTQGGAAQPRHPPAF 120
DB 1165 HLKGTPESSREGTILSSVGLSVLGTBEDEDE-----EEVEYMNRRRHSP-PHPPPS 121
QY 1208 SPAPDNLVYWD-----QDPERGAPRSPFGTPTAENPEYL 1243
DB 1216 SLEELGYEYMDVGSLSASLGSQSCPLHPVPIMPAGITTPBEDYGYM 1263

```

RESULT 10
 ID ERB3 RAT STANDARD; PRT; 1339 AA.
 AC 062739; 062955;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
 DE (C-erbB3).
 GN ERB3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CX NCBI_TaxID=10116;
 RN [1]
 SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RA MEDLINE=96096535; PubMed=8522190;
 RX Hellyer N.J., Kim H.H., Greaves C.H., Sierke S.L., Koland J.G.;
 RT Cloning of the rat ErbB3 cDNA and characterization of the
 RT recombinant protein.";
 RL Gene 165:279-284(1995).
 RN
 RP REVISIONS TO 85: 513 AND 565.
 RA Hellyer N.J., Koland J.G.;
 RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
 RN
 RP SEQUENCE OF 922-1097 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
 RX MEDLINE=97184212; PubMed=9030624;
 RA Carroll S.L., Miller M.L., Froehner P.M., Kim S.S., Corbett J.A.;
 RT Expression of neuroligins and their putative receptors, ErbB2 and
 RT ErbB3, is induced during Wallerian degeneration.";
 RL J. Neurosci. 17:1642-1659(1997).
 CC
 CC - FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTRK.
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC - SUBCELLULAR LOCATION: Type I membrane protein.
 CC - DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
 CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
 CC PHOSPHATIDYLINOSITOL 3-KINASE.
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements)
 CC or send an email to license@ebi.ac.uk).

Query Match	Best Local Similarity	Score	DB 1	Length
5	ALCRWGLTLLP	34.88	Score 2356.5	DB 1: Length 1339
27	AVCP-GTLNGSLVTDADNQVLT	41.18	Pred. No. 3.4e-118	Matches 523; Conservative 169; Mismatches 419; Indels 161; Gaps 34
65	LSFLDIOEVGVYLLANQVRLRLIVAGTGLFEDNYALAVLDNDPLNTPTVT	1028	147545 MW; 0AA5F2402BBDF1E CRC64;	
74	LSFLQWIREVTVGLVVMNEFSVLRNLLVVGTVQDGKFAIFV	1028		
125	GASPGGLRELDRLSLTEILKGVLLIQNPOLCYODTILMIDIFHNKNNQLALTIDTRSR	1028		
126	NSSHALRLQKLTGLTEILSGVYIEKNDLCHMDITDWDIVR	1028		
125	ACHPSPWCKSGSRWGSSEDDCGSLTIVACGG-ARCKPRLTDCCHBOCAAGCTGPRN	1028		
182	NCPDHEVCCKG-RCWGGPDDCQILITKICAPQNGRCFGPNPNOCHCAGCGSGPD	1028		
244	SDCLACLFHNSGICELCPALVTYNTDFESMPNREGRTFGASCTACPYULSTDVG	1028		
241	TDFCRACRFNDSGACVRCRCEPLVYNTLTQLBPNRPTKQYGVGVASCPHNFV-VDQT	1028		
304	SCTLVCPILHNOETVAEDGTORCEKSGPCARVYUGLMEHREVRVAVTSANIQEFAGCK	1028		
300	FCVACAPCPDKKEVD-KGLKWCCECGCLCPKACGEGTSG-SRYCATDSNINIGFVNCIK	1028		
344	IFGSLAFIPESFDDPSNAPRLQPELOVFELEETGLYLSAMPDLSPLDSVQNIQ	1028		
357	ILGNDFLITGLANDPYNHKLPAIDPELVNFRVREITGLNLIQSPRHNANSVSNTLT	1028		
424	VIRRIILHNGAYS-LTLOGIGISWLGRLSRLDELGLALIHNTHTLCSFYVYTWDLDFRN	1028		
417	TIGRSLTYNRGFSLLIKNNLNTSLGRSLKEISAGVVYISANQOCYNNHSLWMTLLNG	1028		
483	PHQALLHTA-NRPDECVAGGLACHQICAGHGWGGGPGQCVNCSQFLGQDEVCRCVL	1028		
477	PSEERLDIKYDRPLGECILAGKVCDBPLCSSGGGCGPQGLSCRYVSBGVVTHCNFL	1028		
542	QGLPREVYNARHCLPCHPCOPONGSVTCGPAPADOCVACAHYKDPFCVACRPSGVXPD	1028		
537	QGEPRFVHEAQCSCRPCLPMBGISTCNCGSSDAACAHFRDPRHCVNSRPHILG	1028		
602	LSYPIPKFPDEBACQPCPINCTHSC-VDDDKGCPRAQRASPLTISVAVGILLV	1028		

RT growth factor receptor";
RL Nature 314:178-180(1985).
RN [8]
RP SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
RX ANALYSIS.
RA MEDLINE=92038942; PubMed=1936959;
RA Raz E., Schejter E.D., Shilo B.Z.;
RT "interallelic complementation among DER/f1b alleles: implications for
RL the mechanism of signal transduction by receptor-tyrosine kinases";
RN Genetics 129:191-201(1991).
RP REVIEW.
RX MEDLINE=97248481; PubMed=9094709;
RA Perrimon N., Perkins L.A.;
RT "there must be 50 ways to rule the signal: the case of the Drosophila
RT EGF receptor";
RL Cell 89:13-16(1997).
CC -1- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
CC MAPK PATHWAY. INVOLVED IN A TRIAD OF DEVELOPMENTAL DECISIONS.
CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
CC ESTABLISHMENT OF VENTRAL CELL FATES. MAINTENANCE OF AMNIOSEROSA
CC AND VENTRAL NEURODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
CC CUTICLE.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
CC PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND
CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: UR-QUITOUSLY EXPRESSED IN EMBRYOS, IN LARVAE,
CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
CC AND THORACIC AND ABDOMINAL GANGLIA.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
CC or send an email to license@ebi-sib.ch).
CC -----
DR EMBL AF058754 AAC08536.1 -
DR EMBL AF058755 AAC08536.1 JOINED.
DR EMBL AF058754 AAC08535.1 -
DR EMBL AF058752 AAC08535.1 JOINED.
DR EMBL K03054 AAAS1462.1 -
DR EMBL K03417 AAAS1460.1 -
DR EMBL K03416 AAAS0965.1 -
DR EMBL K03418 AAAS1461.1 -
DR EMBL AF109077 AAD26134.1 -
DR EMBL AF109078 AAD26132.1 -
DR EMBL AF109082 AAD26132.1 JOINED.
DR EMBL AF109078 AAD26133.1 -
DR EMBL AF109084 AAD26133.1 JOINED.
DR EMBL AF109079 AAD26130.1 -
DR EMBL AF109081 AAD26130.1 JOINED.
DR EMBL AF109079 AAD26131.1 -
DR EMBL AF109083 AAD26131.1 JOINED.
DR EMBL AF109083 AAD26135.1 -
DR EMBL AE003454 AAF46732.1 -
DR EMBL X02293 CAA26157.1 -
DR EMBL X78920 CAA55523.1 -
DR EMBL X78918 CAA55521.1 -

DR EMBL X78919 CAA55522.1 -
DR PIR A00640 GQFE.
DR HSSP E11362 IFK.
DR FLYBASE FBgn0003731 Egfr.
DR InterPro: IPR000494 EGFR_L_domain.
DR InterPro: IPR000719 Euk_Pkinase.
DR InterPro: IPR001274 Furin-like.
DR InterPro: IPR001245 Tyr_Pkinase.
DR Pfam: PF00069 Pkinase.1.
DR Pfam: PF00757 Furin-like.1.
DR Pfam: PF01030 Recep_L_domain.2.
DR PRINTS: PR00109 TYRKINASE.
DR ProDom: PD000001 Euk_Pkinase.1.
DR SMART: SM00264; FU; 7.
DR SMART: SM00219; TyrcK; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
KW Tyrosine-protein kinase; ATP-binding; signal; Alternative splicing;
KW Developmental protein.
FT SIGNAL 1 30
FT CHAIN 31 1426 POTENTIAL.
FT DOMAIN 31 868 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT TRANSMEM 869 889 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 890 1426 POTENTIAL.
FT DOMAIN 938 1198 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 944 952 PROTEIN_KINASE.
FT BINDING 971 971 ATP (BY SIMILARITY).
FT ACT_SITE 1063 1063 ATP (BY SIMILARITY).
FT MOD_RES 902 902 BY SIMILARITY.
FT PHOSPHORYLATION (BY PKC) (BY SIMILARITY).

Query Match 28.4%; Score 1927; DB 1; Length 1426;
Best Local Similarity 32.9%; Pred. No. 17e-95;
Matches 460; Conservative 179; Mismatches 421; Indels 338; Gaps 41

QY 44 YQCGVQVQGNLELYLPT-NASLFLDIOEVOGYVLIANOVRYVPLQRLVIRGTOLF 102
DB 126 YTNCTYVDGNKLTMLPBNENIDSLFDINIEVGYILISHVDYKVFPPKLTQIRGRTLF 185
QY 103 -----EDVYALAVDNDGDPINNTPTVVGASPGGLRELQSLSTLEIKKGVLIQRNPQLCV 157
DB 186 SLVSEEEKYALFV-----TYSKYTLIEDLDVLDVNGVGFNNYNLCH 229
QY 158 QDTLMKDIFKNNQALTLIDTRNSAPCHPCSPMCKSGRSCWSESDCGLTRTVCAQ 217
DB 230 MRLTQSEIIVNGTDAYVNDFTAPERECPKCHESCTHG-CWGEGRPKNGCFSLKTCSPQ 288
QY 218 CA--RCKGPIPTDCHEQCAAGCTGPGHSDCLCLFHNHSGICELHCPALVTVNTQTFES 275
DB 289 CAGRCCTGPKPRRECCHLFCAGGCTGTPQKDCIACKPFDEAVSKECCPPRYKYNPTTVLV 348
QY 276 MPNEGRYTFGASCVTAPYNYLSTDVSGCTVCPHNOEVTAEQSTORCEKSKPCARV 335
DB 349 ETNBEKGVAAGATCVKECP-GHLIRDGACVRSQPDQKMDKGE-----CVPNGCPCPT 402
QY 336 CYGLGMEHLREVAVTSANTQEFAGCKKIGSLAFPESDG--DPASTNA-----PLQ 387
DB 403 CPGVTVLH-----AGNIDSRNCTVIDGNRIIDQTFSGQDVAVYTWGPRYIPLD 454
QY 388 PEQLQVETLEITGYLYISAMPDSJPDLSVFONLQVIRGRIAHNGAY-SLTQGLGISM 446
DB 455 PEREVSTVKEITGYLINIGTHPQPNLSYFNNLETIRQRLQMEFMALAIKSLIS 514
QY 447 LGLASRELQSGALAHNHTLCPVHTVPMWQCFRNPQALHTNANRPEDECVSGGLACH 506
DB 515 LEMENLQISSGSVQIHNNDLCVYSNIRFPAQKPEQKVMYNNIRADLCERKNGTICS 574
QY 507 QLCARHGCMGPGFPQCVNCSQFARGQECVEECVVLQCLPREVYNAHCHLCPRECPQNG 566
DB 575 DQCNEDCGKAGAGTDQCLTCGNFNFNGTCTIADGCVISNAK--FDRNTCKICPECR----- 628
QY 567 SVTCGPEADQCAACHYKDPFCVRCAP-----SGVK----- 599

```

Db      629  --TCAGAGADHCQECVHVRDQHCVCSECPKXKYNDRGVCRECHATCDGCTGPKDTTIGIGA 686
Qy      600  -----PDLSTYMPIMKF--PDEEGACQP-----619
Db      687  CTTGNLATIINDATVKGCLMDKDPD-GY--PMEYVHPQGSGLKLAGRAVCRKHPL 743
Qy      620  CFI-----NCH-----626
Db      744  CELCTNYGHEOVCSKCTHYKRECECECPADHYTDEQRECFQRHPECGCTGPGADD 803
Qy      627  --SC-----VDLDDRG-----CPAEOR-----ASPLTS-----647
Db      804  CKSCNFKLPDANEGPYVNSTMPCSTCKCLERHNYQTALGPYCAASPPSSKITA 863
Qy      648  -----TSAVGVILLVAVGVFGILIKRQOKIRKTY--MRLLQETELVEPLTPS 697
Db      864  NIDVMIFITIGAVLVPTICILCVV--TYICRQKQAKKETVKTMLSCGEDSEPLRPS 921
Qy      698  GAMPQAGMRILKETELRKVAVLGSAGFGTYKGIWIPDGENYKIPAVIAYLENTSPKA 757
Db      922  NIGANLCKLRIVKDELKRGVLMGAFRGYKGVWPEGENVKIPAIKELIKSTGAS 981
Qy      758  NKEILDEAYVAVGSPVYSRLIGICTSTVQVLTQMPYGCCLLDHYENRGRGLSGDIL 817
Db      982  SEEFLEAYIMASEEHVYVLLKTLAVCWSSQWMLITGLMPGCLLDYRNNRDKIGSKALL 1041
Qy      818  NMCMQIAGMSYLEDVRLVHMDLAARVLYKSPHVKYITOPGLARLLDIDETVHADGK 877
Db      1042  NMSTQIAGMSYLEERKLVHDLAARVLYQTPSLVKITDFGLAKLISPSNRYKKAQK 1101
Qy      878  VPIKMALESILRRRFTHQSDVMSYGVTVWELMFPKAPYDGI-PAREIPDLLEKGERLPQ 937
Db      1102  MPKMLAFECIRNNVFTFSKSDVMAFGYTIWELLTFGRPHENIPADIDPLIVGILKEQ 1161
Qy      938  PPICTIDVYIMVCKMIDSECRPFELVSESRVADPQRFVIONEDLG--PASPDL 995
Db      1162  PEICSLDIYCTILSCWMLDAMRPFTFQKLTTFEAFARDGRYLAIPGKFTLRPA-----1217
Qy      996  STFYSLEDD--DMGDVDAEYVLPQGGFPGDPAPAGAGWVHRRSSSTRSGGD 1052
Db      1218  ---YTSQDEKLILKLAFTTDGSAIKKPDYIQPKAFPPS-----HRTDT-----1262
Qy      1053  LTLGLEPSEEBAP-----RSLPAFSEAGSDYVFDG--DLGMAAKGQLSLPTHPSPIL 1103
Db      1263  -----DEMPYLNRVCKDPSKNSSTGDDERDSAREVGVGNLR-----1300
Qy      1104  QRYSEDPVPLPSTDGVAAPLTSPQPEYVQDPVAPQPSRREGPLPAAPDAGATLER 1163
Db      1301  -----LDLPEVDEEDILMP--TCQPGPNNNNMN-----NENQNNMAVGVAAQYM--1343
Qy      1164  AKLSPGKNGVAVKDFAFGAVENPEYL--TPQGAAPQPH-----1202
Db      1344  -----DLIGVPSVDNPEYLLNAQTLGVGESIPITQITIGIVMGPGTMEVK 1390
Qy      1203  -PPAFSP-AFDNLYYND 1218
Db      1391  VPMGSEPTSSDHEYYND 1408

```

```

RESULT 12
ID_ERBB_ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUN-1999 (Rel. 38, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
OS V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
NCBI TaxID=11864;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nilsen T.W., Maroney P.A., Goodwin R.G., Rotman F.M.,
RA Critchenden L.B., Raines M.A., Kung H.-J.,
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726(1985).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INFECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaborative
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M10066; AAA48763.1; ALT_INIT.
CC PIR: A00643; TVCHLV.
CC PIR: B00643; TVPLV.
CC HSSP: P11362; 1FGK.
CC InterPro: IPR000719; Euk_Pkinase.
CC InterPro: IPR001245; Tyr_Pkinase.
CC Pfam: PF00069; Pkinase; 1.
CC PRINTS: PR00109; TYRKINASE.
CC ProDom: PD00001; Euk_Pkinase; 1.
CC SMART: SMO0219; Ty-KC; 1.
CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
CC Trasnferase: Tyrosine-protein kinase; ATP-binding; Oncogene;
CC Glycoprotein; Phosphorylation.
CC DOMAIN 132 399
CC NP_BIND 138 146 ATP (BY SIMILARITY).
CC FT BIND 165 165 ATP (BY SIMILARITY).
CC ACT_SITE 257 257 BY SIMILARITY.
CC SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

```

Query Match 25.8%; Score 1749.5; DB 1; Length 634;

Best Local Similarity 52.3%; Pred. No. 3e-86; Matches 135; Indels 123; Gaps 17

```

Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17
Qy 581 CAHYKPPFCVAVCPGSGVXKPLSYMPIMKFPDEEGACQPCPINTCTHSCVLDLDCGCPAQ 640
Db 3 CAHFIDGPHCVKACPAVGLGENDTL--VMKYADANAVCQCHPNCRTGCGPGLEGCP---58
Qy 641 RASPLTISVAVV--GILLVAVGVGILIKRQOKIRKTYRRLLOETELVEPLTPSCA 699
Db 59 NGSTFPIAGVVGGLCLVVGGLGLVLR--HIVRKRRLRLRLQERELEVPPLPGE 117
Qy 700 MPQAGMRILKETELRKVAVLGSAGFGTYKGIWIPDGENYKIPAVIAYLENTSPKANK 759
Db 118 APQAHRLIKETEPKVKVAVLGSAGFGTYKGIWIPDGENYKIPAVIAYLENTSPKANK 177
Qy 760 EILDEAYVAVGSPVYSRLIGICTSTVQVLTQMPYGCCLLDHYENRGRGLSGDILNW 819
Db 178 EILDEAYVAVGSPVYSRLIGICTSTVQVLTQMPYGCCLLDHYENRGRGLSGDILNW 237
Qy 820 CMQIAGMSYLEDVRLVHMDLAARVLYKSPHVKYITOPGLARLLDIDETVHADGKVP 879
Db 238 CVQIAGMSYLEDVRLVHMDLAARVLYKSPHVKYITOPGLARLLDIDETVHADGKVP 297
Qy 880 IKMMALESILRRRFTHQSDVMSYGVTVWELMFPKAPYDGI-PAREIPDLLEKGERLPQ 939

```

```

Db      298 IKMALESLHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIASISVLEKGERLPQP 357
Qy      940 ICTIDVYMIWYKMWIDSECRPRFRELVEFSRMARDQRFVVIQ-NEDLGPAAPLDSTF 998
Db      358 ICTIDVYMIWYKMWIDADSRKRFRELLAEFSKMRADPRYLVIQGDERRMLPSPDTSKF 417
Qy      999 YRSLEDDMDGLVDAAEYLVLPQCGFFCPDPAPAGAGWVHRHRSSTRSGGDLTLGLE 1058
Db      418 YRTLMEEDEMDYVDADYLVHPHQGF-----NSPST----- 449
Qy      1059 PSEEBAPRSPL-----APSEGAGSVDFGDLGMGAAGLQSLPTHDPSLPORYSEDPVTP 1113
Db      450 -----SRTPLLSLSATSNNSATNCID-----RNGQGHPRVEDSFVGRYSSDPDGN 495
Qy      1114 LPSET--DGYVAAPLTGSPQPEYVNOQDVPPOPSPREGPLPARPAGATLERAKTLPSPGK 1171
Db      496 FLEESIDDGFL-----PAPRYVNO--LMPKRPSTAM----- 526
Qy      1172 NGVVKDVPF-----AFQGAVENTPYLTPOGAAPQHPPPAPSPAFDMLY 1216
Db      527 NQIYNNISLTALSKLPMDSRQNSHSTAVDNPEYL-----NTNQSPPLAKTVFESSPY 578
Qy      1217 WDQ-----DPEE-----RGAPSTFRKGTPTAENPEYIGLDPV 1248
Db      579 WIOGNHGINLNDPDIQODPLFNETKPMGLKVPAPENPEYLRVAAP 625

RESULT 13
ERBB_AVIER STANDARD: PRT: 604 AA.
AC R00515;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbb gene of avian erythroblastosis virus is a member of the src
   gene family.";
RL Cell 35:71-78(1983).
RN (2)
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RA Debuire B., Henry C., Benajissa M., Bisette G., Clavierie J.-M.,
RA Sauls S., Martin P., Stehelin D.;
RT "Sequencing the erba gene of avian erythroblastosis virus reveals a
   new type of oncogene.";
RL Science 224:1456-1459(1984).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
   tyrosine phosphate.
CC -1- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
   ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
   IN CHICKENS.
CC -1- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
   RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL Outstation
   at the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.ebi.ac.uk/announcements
   or send an email to license@ebi.ac.uk).

```

```

DR EMBL: K02006; AAA42394.1; ALT_INT.
DR EMBL: K01216; AAA42400.1; -.
DR PIR: A00644; TYUH.
DR HSSP: P11362; IFGR.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00069; pkinase; 1.
DR ProDom: PD000001; Euk.pkinase; 1.
DR SMART: SM00219; TyrKc_1.
DR PROSITE: PS00107; PROTEIN KINASE, ATP, 1.
DR PROSITE: PS00109; PROTEIN KINASE, TYR, 1.
DR PROSITE: PS00111; PROTEIN KINASE, DOM, 1.
DR TRANSFERASE: Tyrosine-protein kinase; ATP-binding; Oncogene;
   Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP BIND 138 146
FT BINDING 165 165
FT ACT SITE 257 257
FT ACT SITE 257 257
FT CONFLICT 29 29
FT CONFLICT 140 140
FT CONFLICT 146 146
SQ SEQUENCE 604 AA; 67633 MW; 76BECDD06745D609 CRC64;

Query Match 25.1%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 8,6e-84;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 1

Qy 581 CAHYKDPFCVAPRCSPGVKPDLSYVPIKWPDEGACQPCPINCTHSCVDLDDGCPAEQ 641
Db 3 CAHFDGPHCVACACAGVUGENDTL-VAKYDANNAVCLCPNCTRGKGGLEGP--- 58
Qy 641 RASPLTIVSAVV-GILLVVLGVVGLIKRROQKIRKYMRLLOTELYEELTPSGA 69
Db 59 NSKTPSIAGAVGGLICLVVVGAGDIGLYLRR-HYAKRTLRRLQGRLEVEFLTPSGE 11
Qy 700 MNQNAQMLKXETELRKXVUGSGAFGVYVGINPDGENKIRVIAIKVLRNTPRANK 75
Db 118 APNQHRLRLKTEKRVKVGSGFGITVGLMIPGEKXKIVAIKELEASPRANK 17
Qy 760 EILDEAVYAGVSGPYVSRLLGICLTSTVQVLTQIMPGCLLDHVRNRCGLSQDILNM 81
Db 178 EILDEAVYASVDNHNVCRLIGICLTSTVQVLTQIMPGCLLDHVRNRCGLSQDILNM 23
Qy 820 CNOIAKGSYLEDFVLVHRDLAARNVLYKSNHYKTFDGLARLLIDETRYHAGGKVP 87
Db 238 CVOIAKGMVYLEERLVHRDLAARNVLYKTPQHYKTFDGLAKLIGADEKEYHAGGKVP 29
Qy 880 IKMALESLHRRFTHQSVDVMSYGVTVWELMTFGSKPYDGIASISVLEKGERLPQP 93
Db 298 IKMALESLHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIASISVLEKGERLPQP 35
Qy 940 ICTIDVYMIWYKMWIDSECRPRFRELVEFSRMARDQRFVVIQ-NEDLGPAAPLDSTF 99
Db 358 ICTIDVYMIWYKMWIDADSRKRFRELLAEFSKMRADPRYLVIQGDERRMLPSPDTSKF 41
Qy 999 YRSLEDDMDGLVDAAEYLVLPQCGFFCPDPAPAGAGWVHRHRSSTRSGGDLTLGLE 105
Db 418 YRTLMEEDEMDYVDADYLVHPHQGF-----NSPST----- 44
Qy 1059 PSEEBAPRSPL-----APSEGAGSVDFGDLGMGAAGLQSLPTHDPSLPORYSEDPVTP 111
Db 450 -----SRTPLLSLSATSNNSATNCID-----RNGQGHPRVEDSFVGRYSSDPDGN 495
Qy 1114 LPSET--DGYVAAPLTGSPQPEYVNOQDVPPOPSPREGPLPARPAGATLERAKTLPSPGK 117
Db 496 FLEESIDDGFL-----PAPRYVNO--LMPKRPSTAM----- 524
Qy 1172 NGVVKDVPF-----AFQGAVENTPYLTPOGAAPQHPPPAPSPAFDMLY 1216
Db 527 NQIYNNISLTALSKLPMDSRQNSHSTAVDNPEYL-----NTNQSPPLAKTVFESSPY 578
Qy 1213 NLYYWDQDPFERGAPPSTFRKGTPTAENPEY 1248

```

DB 575 SSPWIGSQNHQ-----INLNDPDY 594

RESULT 14

ERBB_AVIEU STANDARD; PRT; 540 AA.

ID ERBB_AVIEU

AC P11273;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).

GN v-ERBB.

OS Avian erythroblastosis virus (strain ts167).

OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.

OX NCBI_TaxID=103898;

RN (1)

RP SEQUENCE FROM N.A.

RX MEDLINE=87064459; PubMed=2878364;

RA Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.,

RT "A single amino acid substitution in v-erbB confers a thermolabile

RT phenotype to ts167 avian erythroblastosis virus-transformed erythroid

RT cells.";

RL Mol. Cell. Biol. 6:1751-1759 (1986).

CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein

CC -I- tyrosine phosphate.

CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; M13179; AAA42401.1; -.

DR PIR; A25231; TVEVER.

DR HSSP; P13362; IFGK.

DR InterPro; IPR000719; Euk_kinase.

DR InterPro; IPR001245; Tyr_kinase.

DR Pfam; PF00069; Pkinase; 1.

DR ProDom; PD000001; Euk_kinase; 1.

DR SMART; SMO0219; Tyrc; 1.

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.

DR Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;

KM Glycoprotein; Phosphorylation.

KM DOMAIN 132 399

FT NP BIND 138 146 ATP (BY SIMILARITY).

FT BINDING 165 165 ATP (BY SIMILARITY).

FT ACT SITE 257 257 BY SIMILARITY.

FT VARIANT 270 270 H->D (IN THERMOLABILE V-ERBB).

SQ SEQUENCE 540 AA; 60412 MW; 5853297AA06B65D CRC64;

Query Match 24.1%; Score 1630; DB 1; Length 540;

Best Local Similarity 55.1%; Pred. No. 5.8e-80;

Matches 341; Conservative 69; Mismatches 119; Indels 90; Gaps 14;

QY 581 CAHYKDPFCVAPRPSGVKRPPLSTMPYWKPPDESGACQPPINCTHSQVLDLDDKGPAPQ 640

DB 3 CAHFIDGHCVKACPAVGLGENDTL-VWKYADANAVQOLCHPCTGCKGPGLEGFP--- 58

QY 641 RASPLTISVAVV-GILLVVLGVVFGILLIKRROKIRKTYTMRLLQETFLVPLTPSGA 699

DB 59 NGSKTPSIAAGVVGGLGLVVGGLGYLYRR-HYRKRTLRLLQERLVEPLTPSGE 117

QY 700 MPNQQRILKETLRKRYVLSGAFCTVYKGLWIPDGENVKIPVAKVLENTSPKANK 759

DB 118 APNQAHRLKETLRFKRYVLSGAFCTVYKGLWIPDGENVKIPVAKVLENTSPKANK 177

QY 760 EILDEAVVMAGVSPYVSRLLIGITLTSTVOLVQWQMYGGLDHWENGRGLSQGLLW 819

DB 178 EILDEAVVMAGVSPNPHVCRLLGLICLTSTVOLTLQMPYGCILDIYREHKONIGSQYLLW 23

QY 820 CMQIAKGSYLEDRVLVYRDIAARNVLYKSNHVKITDFGLARLLDDEFEYHADGKVP 87

DB 238 CVQIAKGNVYLEERHMYARDIAARNVLYKSNHVKITDFGLARLLDDEFEYHADGKVP 29

QY 880 IKMALESILRRFTHGSDVMSYGVATWELMTFGAKPYDGPAPAEIPDLLEKGRLPQP 93

DB 298 IKMALESILRRITTHGSDVMSYGVATWELMTFGSKRYDGPAPAEISVYLEKGRLPQP 35

QY 940 ICTIDVYIMYKCMWIDSECRPRREIVSEFSRMAPDPQRFVVIQ-NEDIGPASPDLSTF 99

DB 358 ICTIDVYIMYKCMWIDSECRPRREIVSEFSRMAPDPQRFVVIQ-NEIGPASPDLSTF 41

QY 999 YRSLEDDMDGLDVAEYVLPQGFPCPPAPAGAGVHHRHSSSTRGGGLTLGLE 10

DB 418 YRTLMEEEDMEDIDVADSYLVPHGFF-----NSPT 44

QY 1059 PSEEPAPRSP-----APSEAGSDVFDGDLGMAKAGLSLPTHDSPLOQRYSEDTVP 11

DB 450 -----SRPLSLSLATSNNSATNCIDRNG-----H----- 47

QY 1114 LPSEIDGYVAALTSPPQPEYVNOQDVNPQPSRREGPLPAPRPAAGT-LERAKTISPKN 11

DB 477 -PVREDGFL-----PAPEYVNO--LMPKRPSTAMVQNIYVISTAIKSLPDSRYQN 52

QY 1173 GVYKDVFAFGAVENPEYL 1191

DB 528 -----SHSTAVNPEYL 539

RESULT 15

EGFR_CHICK STANDARD; PRT; 703 AA.

ID EGFR_CHICK

AC P13387;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-JAN-1990 (Rel. 13, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER

DE Fragment).

GN EGFR.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae;

OC Gallus.

OX NCBI_TaxID=9031;

RN (1)

RP SEQUENCE FROM N.A.

RX MEDLINE=88261272; PubMed=3260329.

RA Lax I., Johnson A., Hawk R., Sap J., Bellot F., Winkler M.,

RA Ullrich A., Vennström B., Schlessinger J., Givol D.,

RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,

RT expression in mouse cells, and differential binding of EGF and

RT transforming growth factor alpha.";

RL Mol. Cell. Biol. 8:1970-1978 (1988).

CC -I- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,

CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND

CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).

CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein

CC tyrosine phosphate.

CC -I- SUBCELLULAR LOCATION: Type I membrane protein.

CC -I- MISCELLANEOUS: Binding of EGF to the receptor leads to

CC dimerization, internalization of the EGF-receptor complex,

CC induction of the tyrosine kinase activity, stimulation of cell DNA

CC synthesis, and cell proliferation.

CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch.

```
CC -----
DR EMBL: M20386; AAA48760.1;
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR SMART: SMO0261; FU; 4.
DR PROSITE: PS0107; PROTEIN_KINASE_ATP; PARTIAL.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; PARTIAL.
DR PROSITE: PS00111; PROTEIN_KINASE_DOM; PARTIAL.
DR Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT STGNL 1 30
FT CHAIN 1 30
FT DOMAIN 31 654
FT TRANSMEM 31 654
FT DOMAIN 655 667
FT DISULFID 197 206
FT DISULFID 201 214
FT DISULFID 222 230
FT DISULFID 226 238
FT DISULFID 239 247
FT DISULFID 243 255
FT DISULFID 258 267
FT DISULFID 271 298
FT DISULFID 302 314
FT DISULFID 318 333
FT DISULFID 336 340
FT DISULFID 313 322
FT DISULFID 517 530
FT DISULFID 533 542
FT DISULFID 546 562
FT DISULFID 565 581
FT DISULFID 569 589
FT DISULFID 592 601
FT DISULFID 605 627
FT DISULFID 630 638
FT DISULFID 634 646
FT CARBOHYD 134 134
FT CARBOHYD 190 190
FT CARBOHYD 200 200
FT CARBOHYD 359 359
FT CARBOHYD 368 368
FT CARBOHYD 420 420
FT CARBOHYD 573 573
FT CARBOHYD 578 578
FT CARBOHYD 613 613
FT CARBOHYD 633 633
FT CARBOHYD 648 648
FT NON_TER 703
SQ SEQUENCE 703 AA; 77427 MW; AFF2DE11B735A690 CRC64;
```

Query Match 23.4%; Score 1586; DB 1; Length 703;
 Best Local Similarity 44.1%; Pred. No. 1.7e-77;
 Matches 312; Conservative 113; Mismatches 247; Indels 36; Gaps 14;

```
QY 8 RMGLLALLPPGAA-----STQVC-----TQYIKANSKFIGITELLVYGGCOVQGNL 54
DB 13 RGAALVLLLLGLVALCSAVEEKVCQGTNNKLTOLGHEVDHFTSL-QRMVNNCEVVLSTNL 71
QY 55 ELTYLPTNASLSFLQDIEVQGVLIANQVROVPLQRLIVRGTOLEPDNYALAVLDNG 114
DB 72 EITVEHNRDLIFLKTIOGVAGVLIALNMVDVILENLQIIRGVNLYDNSFALAVLSNY 131
QY 115 DPLANTTPVTGASPGGLRLQLRLSLTEILKGGVLIQRNPQLCYODTIIMKDIIFKNNOLA 174
DB 132 H-MNKTQ-----GLRELPMKRLSELINGVKISNNPKLCNMDVILMNDIIDSRR-P 181
QY 175 LFLID-TNRSRACHPCSPMKCKGRCMGSSSEDCOSLTRVCAGGA-RCKGLPTDCHE 232
```

```
DB 182 LTVLDPASVLSGPCPKCHPCTEDHCHGAGBQNCQTLTVICAGQCCSGRCRGKVPSPDCCHN 24
QY 233 QCAAGCTGPKHSDCLACHFNHSGICEILHCPALVYNTDTESMNPREGRTYFGASCYTA 29
DB 242 QCAAGCTGRESDDCLACRFBDATCKDTCPLVLYNPTTYQMDVNPBGKYSFGATCYRE 30
QY 293 CPNNYISTDVSGCTLVCPILHNOEVAIEGTOQCEKSCSPCARVVCYGLMEHLREAVATS 35
DB 302 CPNNYVTHGSCVRSQNTDTYEV-EENGVRCKKCDGLCSKVCNGIGIGELKGLISTNA 36
QY 353 ANIQEPAGCKXIFGSLAPFESFGDDPASNTAPLOPEOLQVFEETLEITGYLYISAWPDS 41
DB 361 TNDSFKNCTKINGDVSILPVAFLDAPFKTLPLDPKXLDVFRVKEISGFLLIQAWDN 42
QY 413 LPPLSPQNLQVIRGRILHNGYSITLQGLISWGLSLSELSGLALIHNNHLCVH 47
DB 421 ATDLVAFENLEIIRGRVQHGYSLAVNVLKIQSLGLSLKEISDGLAIKNNVLCYAD 48
QY 473 TVPMDQLFENPQALLHTANREDECEVGEGLACHQLCARGCHWGPGPYQVNCQFLRGQ 53
DB 481 TMMRSLFATQSQKTKIIONRYKNCCTADRHVCOPLCSDVCGMGGRPHCSRRFSRQK 54
QY 533 ECYEGCVLQGLPREYVNAHCLPCHPCQPNG---SVTCFGRPADQCVACAHYKDPF 58
DB 541 ECYKQCNIIQGEPRERFERSKCLPCHSECLVQNSTAVNTTCSGPGPDHCKMAHFIDGP 60
QY 590 CVARCPGVPPLSYMPTMKPPDEGACOPINCTHSQVLDKKGCPAQRASPLTSIV 64
DB 601 CVKACPAGVLENDTL-VKXADANAVALQHPNCTRCKGPGLEGCP--NGSKTBSIA 65
QY 650 SAVV-GILLVVVLGVVFGILIKRQOKIRKXTMRLLGETELVEPLTP 696
DB 657 AGVAGGLCLVVVVGIGIYLR-R-HIVRKRTLRLLGERELVEPLTP 703
```

Search completed: July 22, 2003, 08:43:43
 Job time : 20.1815 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.1263 Seconds
(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-5-25-12

Perfect score: 6774

Sequence: 1 METALCRWGLLALPPGA.....TFKGTPTAENPEYLGDVPEV 1249

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_protent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6161	91.0	1259	6	018735 canis fam1
2	3141	46.4	1209	1	090X70
3	3112	45.9	1210	11	09EP98
4	2730	40.3	1165	13	09Y440
5	2729.5	40.3	1137	13	09W6F6
6	2292	33.8	1328	5	P79754
7	2000.5	29.5	1433	3	09BIR9
8	1740.5	25.7	419	4	09UK79
9	1739	25.7	367	11	08R2X1
10	1720	25.4	729	15	086712
11	1718	25.4	567	15	086714
12	1697.5	25.1	412	4	08WYU0
13	1653.5	24.4	962	15	064895
14	1645	24.4	545	15	085468
15	1503.5	22.2	655	11	09WYF5
16	1487.5	22.0	643	11	09ERV6

17	1282	18.9	1193	5	09Y1X8	09Y1X8 ephydaria f
18	1188.5	17.5	1368	5	023821	023821 caenothabdi
19	1182	17.4	1717	5	026566	026566 schistosoma
20	1117	16.5	527	13	090836	090836 gallus gall
21	998.5	14.7	478	11	09ES80	09ES80 rattus norv
22	933.5	13.8	599	13	09PSH2	09PSH2 gallus gall
23	906	13.4	165	4	014256	014256 homo sapien
24	887	13.1	176	11	0923V5	0923V5 rattus norv
25	806.5	11.9	346	13	P11776	P11776 xiphophorus
26	778	11.5	435	5	08SZW1	08SZW1 drosophila
27	754.5	11.1	311	13	09P162	09P162 xiphophorus
28	752	11.1	1362	13	09PVZ4	09PVZ4 xenopus lae
29	738	10.9	1671	5	09NVU5	09NVU5 biophthalari
30	723	10.7	149	6	09BG66	09BG66 oryctolagus
31	716.5	10.6	331	4	09BUD7	09BUD7 homo sapien
32	705	10.4	1418	13	093457	093457 scophthalm
33	696.5	10.3	1368	13	08UW85	08UW85 paratichthy
34	675.5	10.0	1369	13	08UW86	08UW86 paratichthy
35	670	9.9	1358	13	073798	073798 xenopus lae
36	658	9.7	1412	5	08UW84	08UW84 bombyx mori
37	657	9.6	1472	5	09U5A8	09U5A8 paratichthy
38	651.5	9.6	1418	13	08UW83	08UW83 paratichthy
39	645	9.5	1245	13	09YGH8	09YGH8 scophthalm
40	643	9.5	1371	11	09OCW4	09OCW4 drosophila
41	617	9.1	2144	5	09VD54	09VD54 rattus norv
42	595	8.8	987	11	091YU0	091YU0 mus musculu
43	593	8.8	935	4	096LJ5	096LJ5 homo sapien
44	592	8.7	987	11	099WR2	099WR2 mus musculu
45	587.5	8.7	1036	4	007912	007912 homo sapien

ALIGNMENTS

RESULT 1
018735
AC 018735 PRELIMINARY: PRT: 1259 AA.
DT 01-VAN-1998 (TRENBLREL. 05, Created)
DT 01-VAN-1998 (TRENBLREL. 05, Last sequence update)
DT 01-VAN-2002 (TRENBLREL. 21, Last annotation update)
DE ExB-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Carnivora; Fissipedia; Canidae; Canis.
OK NCBI_taxid=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RU EMBL; AB008451; BAA23127.1; -
DR HSSP; P11562; IFGK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_DKinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_Kinase.
DR InterPro; IPR004019; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR Pfam; PF02757; YLP_2.
DR ProDom; PD006001; Euk_DKinase; 1.
DR SMART; SMO0251; FU; 3.
DR SMART; SMO0219; TyKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Tyrosine-kinase.
SQ SEQUENCE 1259 AA; 137969 MW; E3364D45C4ACD46 CRC64;

Query Match 91.0%; Score 6161; DB 6; Length 1259;
 Best Local Similarity 90.4%; Pred. No. 0;
 Matches 1139; Conservative 43; Mismatches 66; Indels 12; Gaps 3;

QY 1 MELALCRWGLLIALLPGAASTVCT-----QYIKANSKFIGITELLXGCGVAGNL 54
 DB 1 MELAAWCRWGLLIALLPGAASTVCTDMKRLPASPEITHLMLRHLXGCGVAGNL 60

QY 55 ELTYLPTNALSFLDIOIEVGVYLIANOVQVQLRIRVGRQLEFEDYALAVLDNG 114
 DB 61 ELTYLPANASISFLDIOIEVGVYLIANSOVQVQLRIRVGRQLEFEDYALAVLDNG 120

QY 115 DPLNNTVPTGASPGGLRELOLRSLTEILKGGVLIQRNPOLCYQDTILMKDIFKNNOLA 174
 DB 121 DPLEGGIPAPGAAOGGLRELQLRSLTEILKGGVLIQRNPOLCYQDTILMKDIFKNNOLA 180

QY 175 LTLIDNRSRACHPSGPKCKSGKSGESSEDCOSLTRVCGAGGACGKGPPTCCHEQC 234
 DB 181 LTLIDNRSACRPSGPKCKSGKSGESSEDCOSLTRVCGAGGACGKGPPTCCHEQC 240

QY 235 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYTTDFEAMPNREGRTFGASCVTACP 294
 DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYTTDFEAMPNREGRTFGASCVTACP 300

QY 295 YNYLSTDVGSCTLYGPIHNOEYTAEDGTQCKGCKSPCARVCTGIGNEHLREVAATVSAN 354
 DB 301 YNYLSTDVGSCTLYGPIHNOEYTAEDGTQCKGCKSPCARVCTGIGNEHLREVAATVSAN 360

QY 355 IOEFAGCKKIFGSLAFLPESFDPDASNTAPLOPEOLVFEFTLEITGYLYISAMPSLP 414
 DB 361 IOEFAGCKKIFGSLAFLPESFDPDASNTAPLOPEOLVFEFTLEITGYLYISAMPSLP 420

QY 415 DLSVFQNLQVIFGRLLHNGANSLTLOGIGISWLGIRSLRELGSLALIHNTHLCEVHTV 474
 DB 421 NLSVFQNLQVIFGRLLHNGANSLTLOGIGISWLGIRSLRELGSLALIHNTHLCEVHTV 480

QY 475 PWDOLFRRPHQALLTANRPEDECEVGEGLACHQICARHCQGPRTQVNCOSQLRQEC 534
 DB 481 PWDOLFRRPHQALLTANRPEDECEVGEGLACHQICARHCQGPRTQVNCOSQLRQEC 539

QY 535 VEECRVLOGLPREYVNAHCLPCHCECQPNQSVTCFQPEADQCVACAHYKDPFCVARC 594
 DB 540 VEECRVLOGLPREYVNAHCLPCHCECQPNQSVTCFQPEADQCVACAHYKDPFCVARC 599

QY 595 PSGVAPDLSTYPIWKEPDEGACOCPCPNCHSCVDLDDKCPAEQASLTSTVSANVG 654
 DB 600 PSGVAPDLSTYPIWKEPDEGACOCPCPNCHSCVDLDDKCPAEQASLTSTVSANVG 659

QY 655 ILLVVLGVFGLIKRRQOKIRKYTMERLLQETELVEPLTPSGAMPNQAMRILKETEL 714
 DB 660 ILLVVLGVFGLIKRRQOKIRKYTMERLLQETELVEPLTPSGAMPNQAMRILKETEL 719

QY 715 RKKVYLGGAGFYKGIWIPDGENVKIPVAIKYLRNTSPKANKELDPAAYMAGVGSF 774
 DB 720 RKKVYLGGAGFYKGIWIPDGENVKIPVAIKYLRNTSPKANKELDPAAYMAGVGSF 779

QY 775 YVSRLLGICLTSTVQLVQLMPYCLLDHVENGRGSGDGLNMCQIAKMSYLEDVR 834
 DB 780 YVSRLLGICLTSTVQLVQLMPYCLLDHVENGRGSGDGLNMCQIAKMSYLEDVR 839

QY 835 LVHRDLAARNLVVSPNHVKITTDGLARLLDIDETEHYADGKVPILKMALESILRRPT 894
 DB 840 LVHRDLAARNLVVSPNHVKITTDGLARLLDIDETEHYADGKVPILKMALESILRRPT 899

QY 895 HQSDVMSYGVTVMLMTFGAKPYGIPAREIPDLLEGEKLPPOPICTIDVYIMYKCM 954
 DB 900 HQSDVMSYGVTVMLMTFGAKPYGIPAREIPDLLEGEKLPPOPICTIDVYIMYKCM 959

QY 955 IDSECRPAFRELVEFSRMAPDQRFVVIQNEDLGPASPLDSTFYSSLDDMDGLVDA 1014
 DB 960 IDSECRPAFRELVEFSRMAPDQRFVVIQNEDLGPASPLDSTFYSSLDDMDGLVDA 1019

QY 1015 EBYLVPOQGFPCPDAPGAGMHHRRSSSTSGGSDLTLGLPEBEERAPSPPLASBG 1074

DB 1020 EBYLVPOQGFPCPEPTPAGTAHRRHSSSTRNGGGLTLGLPSEEPKSPPLASBG 10
 QY 1075 AGSDVFDGDLGGAAGKAGLSLPTHDPSPLORYSEDPYLPPLPSETDGYAPALTCSPPEYV 11
 DB 1080 AGSDVFDGDLGGAAGKAGLSLPSQDPSPLQYSEDPYLPPLPSETDGYAPALTCSPPEYV 11

QY 1135 NQPDVRFQPSRPEEPPLAARPAATLER-----AKTILSPKRGVYKQVAFAGAVENPE 11
 DB 1140 NQPEVWPQPPALAEPPPLPSRPAATLERPTLSLPSKRGVYKQVAFAGAVENPE 11

QY 1190 YLTPQGAAPCPHPAPSPAFNDNLYYWDQPPRGAAPSTFFKGPPTAENPEYGLDVPV 12
 DB 1200 YLAPRGAAPCPHPAPSPAFNDNLYYWDQPPRGAAPSTFFKGPPTAENPEYGLDVPV 12

RESULT 2
 ID Q9QX70 PRELIMINARY; FRT; 1209 AA.
 AC Q9QX70;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Epidermal growth factor receptor.
 GN EGFR.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxId=10116;
 RN (1)
 RP SEQUENCE FROM N. A.
 RC STRAIN=FISHER, TISSUE=LIVER;
 RA MEDLINE=90258888; PubMed=2342465;
 RA Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,
 RA Barp H.S.;
 RT "A truncated, secreted form of the epidermal growth factor receptor is
 RT encoded by an alternatively spliced transcript in normal rat tissue."
 RL Mol. Cell. Biol. 10:2973-2982(1990).
 RN (2)
 RP SEQUENCE FROM N. A.
 RC STRAIN=FISHER, TISSUE=LIVER;
 RA Petch L.A.;
 RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE FROM N. A.
 RC STRAIN=FISHER, TISSUE=LIVER;
 RA Guttridge K., Dawson T.L., Barp H.S.;
 RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
 DR HSEF; R11362; 1FGK.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00063; pkinase; 1.
 DR DR Pfold; PF01030; Reced_L_domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 3.
 DR SMART; SM00219; TYRK; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00101; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR KATP-binding; Receptor; Transferrase; Tyrosine-protein kinase.
 SQ SEQUENCE 1209 AA; 134891 MW; 96FEETFFCCIB7773 CRC64;

Query Match 46.4%; Score 3141; DB 11; Length 1209;
 Best Local Similarity 50.1%; Pred. No. 1,7e-228;
 Matches 636; Conservative 168; Mismatches 360; Indels 106; Gaps 25

QY 3 LAALCRWGLLIALLPGAASTVCTGYIKANSKFIGITELLXGCGVAGNLTYLPTN 62


```

Db      15 LAALCAAGALEKKVCGGTSNR.LTQLGTEDEHFLSL-QRMFNCEVVLGNLEITYVOEN 73
Qy      63 ASLSFLADIDIEVOGVLLIAHQVQVPLORLRIVRGTLQFEDNYALAVDNGPLNNTTP 122
Db      74 YDLSEFLKTIQEVAGVLLIANTVRIPLNLCIIRGALIENTYALAVSN----- 124
Qy      123 VTGASPGSLRELQRLSLTEILKGGVLLIQRPOLCYQDTILMKDIFKNNQALATLIDTNR 182
Db      125 -YGTNKTGLRELPRNRNIOETILIGAVRFSNNPILCNMETIOMRDIY-QDVELSNMSDVOR 182
Qy      183 S-RACHPCSPKCSKRCMGSESSEPCQSLTRTVACAGGA-RCKGPLPDDCCECAAGCTG 240
Db      183 HLTGCPKDCPCSPGSCMGREECQKLTIKIICAQOCSRRCRSPSDCCNCAAGCTG 242
Qy      241 PKHSDCLACLHFNHSGICEHCPALVTYNTDPESMPEGRYTFGASCVTACPYAVLST 300
Db      243 PRESDCLVCHFRDEACIKDCTCPPLMLYNPTTYQMDVNPBGKXSFQATCYKCKPRRYVVT 302
Qy      301 DVGSECLVCPRLHNEVYAEDEGTQCEKSKPCARVCYGLCKEHLREVAVTSANIOEPAG 360
Db      303 DHGSCVRAAGPDYIEV-EEDGVSKCKKCDPCRCVKVCGIGEIFKDTLSINATINIGHFYX 361
Qy      361 CKKIFGSLAPLPESFPDDPASNTAPLQPEQLOVETLEETGYLYISAMPDSLPIQSVFQ 420
Db      362 CTAISGDHLILPVAFKDSFTRTPLDPRLEILUKYTEITGSLLIQAMPENMTDLHAE 421
Qy      421 NLQVIRGRILHNGAVSLTLOGLGIWGLRLSRLBELSGSLAIHHNTHLCFVATVPDQLE 480
Db      422 NLEIRGRKHQGFSLAVVGLNITSIGRLSLKEISGDIIVISGNNTLCYANTINMKLIF 481
Qy      481 RNPQALLHTANREDESGEGLACHQLCARHGCMGPRQCNCSQPLRGQCEVECV 540
Db      482 GTPNQKIKIMNRREKCKXATNHVCNPLSEGGCBPPLDCSCQNVSRGRGCDKCKI 541
Qy      541 LQGLPREVYNAHCLPCHPECQPNQSGSVTCFGEADQCAVACAHYXOPFCVARSQVAP 600
Db      542 LEGEPREFEVSECIQHPCECLPQTMNITCTGRGPNCIKCAHYVDPHCVTCPBSGIMG 601
Qy      601 DLSYMPIKXPPDEGACQPCPINCITHSQVLDKCGPAEDRASPLTSIVSAVGLLVY 659
Db      602 ENNTL-VMKRADANNVCHLCHANCCTGCAGPLKGC--QOPEGPXKPSIATGIQVGLLFI 658
Qy      660 VLGVVFGI-LIKEROQKIRKYRWRLLOETELVEPLTPSGAMPNQAKMILKTELRLKX 718
Db      659 VV-VALGIGLFMRRLQVLRKTRILRLLOERELVEPLTPSGEARNQHLILKTEFEKIK 717
Qy      719 VLGSAGFTYKGIWTPDGENVKI PVAIKVLRNTPSPKANKEILDEAVYVAGVSGPYSR 778
Db      718 VLGSAGFTYKGIWTPDGENVKI PVAIKVLRNTPSPKANKEILDEAVYVAGVDPHYCR 777
Qy      779 ILGICLTSTVQLTQLMPPYGLLDHYRENGRLGSGODLNMCMQIAKMSYLEDRLVHR 838
Db      778 ILGICLTSTVQLTQLMPPYGLLDHYREKDNIGSYQLNMCCQIAKMGNYLEDRLVHR 837
Qy      839 DLAAARNVLKSPNVHKTIDFGLARLLDIDETEHADGKVPKIMWALBESILRRRFTHSD 898
Db      838 DLAAARNVLKTPOHVITIDFGLAKLLGAEEKEYHAEGKVPKIMWALBESILHRIYTHSD 897
Qy      899 VMSYGVTVWELMFEAGKPYDGIAREIPDLLEKGEELPPRPICTIDYVIMKCMKNIASE 958
Db      898 VMSYGVTVWELMFEAGKPYDGIAREIPDLLEKGEELPPRPICTIDYVIMKCMKNIAD 957
Qy      959 CRPRFRELVSERMARADPORFVVIQ-NEDLGPASPLDSTFRSLIEDDDMDGLVDAEEY 1017
Db      958 SRPKFRELILFESKMARDPQRYVILVIGDERMHLPSPTDGNFRAIMBEEDEMDVDADAEY 1017
Qy      1018 LVPOQGFPCDPAFAGAGCMVHHRHSSSTRSGGDLTLGLBSESEAPRSPLABEGAGS 1077
Db      1018 LIPQGGF-----NSPST-----SRPILSLISANS 1043
Qy      1078 DVEDGDLGMGAAGKGLSLPTHPSPLQRYSEDPTVLPSET--DGVVAPLTCSPPPEVYN 1135
Db      1044 N-----SSTVACINRNSCRVKEDAFIQRRSSPTSVTLTBDNIDTFL-----PYPEYIN 1093

```

```

Qy      1136 QPDYRPPQPSRREGPLPAARPAQTILERAKTSPGKGVKDYFAFGAVENPEYL-TQ 113
Db      1094 Q-SYVKKPPAGSVQVPHNPPHLP-----APGDLHYQN--PSNVAVSNEYLNTAQ 114
Qy      1195 GGAAPQHPHPAPFAPADNLVYDQ-----DP-----PERGAPPSTFGTPTAE 123
Db      1143 -----PTCLSSGFDSSALMTQKSHQMSLDNDYQODFFPKAKPNIGIFKG-PTAE 119
Qy      1239 NPEYLGLDVP 1248
Db      1193 NAEYLRAVAP 1202

```

```

RESULT 3
Q9EP98 PRELIMINARY; PRT; 1210 AA.
ID Q9EP98;
AC Q9EP98;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor isoform 1.
CN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CH/101, 129/SVJ, AND 129/SVEV7AC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Sinclair C.S., Pearshall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Mathle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egrf transcripts encoding truncated receptor
RT isoforms."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Schenl C., Pearshall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Mathle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egrf transcripts encoding truncated receptor
RT isoforms."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

```

KW ATP-binding; Receptor; Transferase.
 SEQ SEQUENCE 1210 AA; 134840 MW; 62CD021C9DE32E18 CRC64;
 Query Match 45.9%; Score 1112; DB 11; Length 1210;
 Best Local Similarity 49.5%; Pred. No. 2.7e-226;
 Matches 630; Conservative 169; Mismatches 356; Indels 118; Gaps 25;

11 LLLALLPQAA--STQVC-----TOYKANSKFIGITELLVQGCQVQGNLELTLP 61
 14 LITLCAAGALEEKVCQGSNRLTQGTGFEDHFLSL-QRMYNCCVVLGNLEITVQR 72
 62 NASLSFLQDIQVQGVYLIANQVQVLPQRLIVRTQTLFEDNYALAVLDNGDPLNNT 121
 73 NYDISFLKTIQVQGVYLIANQVQVLPQRLIVRTQTLFEDNYALAVLDNGDPLNNT 124
 122 PVTGASPGLELQRLSTELLKGVLIQIRNPQICQOPTIMKDI----PHKNNQALATL 177
 125 --YGTNRGTGLRLPRNQLQELIGAVRSNNPILCNMDITQMRITVQVMSKMSDL-- 180
 178 IDNRSRACHPCSPCKGSRCKGSESSDCCSLTRTVACGCA-RCKGPLPTDCCHEGCA 236
 181 --QSHPSSCPCKDPCPGKSCMGGEENCQKLTXTIICAQCQSHRCGRSPSDCHNQCAA 238
 237 GCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMNPBGRYTPGASCTACPN 286
 239 GCTGPRESDCLVCKQFOEATCKOTCPPLMYNTTQMDVNPBGKYSFGATCVKCKCPRN 298
 297 YLSTDVSGCTIVCPILHNGEVTAEQGTORCEKSPKPCARVCYGLMEHLREAVATSANTQ 356
 299 YVVDHSGCVACGPDYEV--EEDGIRKCKKCDGPCRCKVNGIGIGIFKDTLSINATNIK 357
 357 EFACGCKIFGSLAFIPESFDDGPASNTAPLOPELQVFEILLETITGLYISAPDLSPLD 416
 358 HFKYCTAISGLHLIPVAFKDSSTRPPLDPRLELTIKVETLTKTGLLIQAAPDMWTLL 417
 417 SVFQNLQVIRRIILHNGAYSLTLQGLISWLGSRSLRSLSGSLAIHNTLHLCFVATVP 476
 418 HAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLRSLKEISDGVITISGNNNLTANTIM 477
 477 DQLFNPHQALLHMANRPEDECVCEGLACHQLCAKRGCMRGPPLQCCYNGCQPLRGECE 536
 478 KKLFGTPNQTKIMNNRAEKDCKAVNHVCMPLCSSEGCWPEPRDCCSQQNSRGECVE 537
 537 ECRVLOGLLPREYVNAARHCLPCHPECCOPONGSVTFPGSEACQVACAYKCPRECVARCS 596
 538 KCNLEGEPRFVENSECICQHPCLPQANNITGTGCPNCCQCHAYIDGPHCVTCTCA 597
 597 GVKPDLSTMPWKPPDEGACQPPINCTHSCVDLDDKGGCPABQASPLTSTIVSAVAGIL 656
 598 GIMGENNTL-VWKYADANNVCHLCHANCTYGACGPGIQCCEVMPSPGKIPSIATGIVGL 656
 657 LVVVLGVVFGI-LTKRRQOKIRKXTMRLLQETELVPLTPSGAMQNOQMRILKTELR 715
 657 LPIIV-VALGIGLPMRRHRYKKTLLRLQERLVEPLTPSEAPRQAHRLKTEKER 715
 716 KKVVLGSGAFSTYKGIWIPDGENVKIPVAIKVLRNTSKRANKELIDEAYYAGVGSFY 775
 716 KIKVLGSGAFSTYKGIWIPDGENVKIPVAIKVLRNTSKRANKELIDEAYYAVASVDNH 775
 776 VSRLLGICLTSTVOLVQLMPYGLLDHVENRGLSGSODLLMWQIMKMSYLEDVYL 835
 776 VCRLLGICLTSTVOLVQLMPYGLLDHVENRGLSGSODLLMWQIMKMSYLEDVYL 835
 836 VHRDLAARNVLVSKPNHVKITDFGLARLDIDETEHADQKVPIMKMALESILRRFTH 895
 836 VHRDLAARNVLVSKPNHVKITDFGLARLDIDETEHADQKVPIMKMALESILRRFTH 895
 896 OSDVMSGVYVWELMTGAKPYGIGIPARELPDLLENGERLPQPPCTIVYVIMYKCMWI 955
 896 OSDVMSGVYVWELMTGAKPYGIGIPARELPDLLENGERLPQPPCTIVYVIMYKCMWI 955
 956 DSECRPRFRELVSFSSMARDPQRFVVIQ-NEDLGPASPLDSTFFYSLDEDMGLVDYA 1014

DB DADSRKPRELILEFSKVARDPQRYVLIQGDERNHLPSPTDSNPFYALMDEEDMEDVYDA 101
 QY EBYLVPOQGFCTCPDPAFAGAGMVAHHRSSSTSGGSDLTLLGPSEBEAPRSLAPSG 107
 DB DEYLVPOQGF-----NSPST-----SRTPLLISLS 104
 QY AGSDVFDGDLGMGAAGKGLQSLPTHDPFLQRYSEDPVPLPSET--DGYVAPLTCPOPE 113
 DB ATSN-----NSTVAICINNGSGRVYEDAFLORYSDPFGAVTEIDNIDAF-----PVPE 105
 QY YVQ-SYKPRPAGSVQNPVHNGPLHP-----AGRDLYQN--PHSNVAGPEYLN 114
 DB TPQGAAPQHPHPAPFAPDNLVYMQ-----DP-----PERGAPSTFKGT 123
 QY 1191 TAQ-----PCLSSGFNSPALMQLQKSHQSLNDPQDFFPKETKPNGITKPG-K 119
 DB 1236 TAENPEYLGIDVP 1248
 DB 1191 TAENAEYLRVAPF 1203

RESULT 4
 Q9YH40 PRELIMINARY; PRT; 1165 AA.
 AC Q9YH40;
 DT 01-MAY-1999 (Tremblrel.10, Created)
 DT 01-OCT-2000 (Tremblrel.15, Last sequence update)
 DT 01-JUN-2002 (Tremblrel.21, Last annotation update)
 DE Receptor tyrosine kinase proto-oncogene.
 GN XMRK.
 OS Xiphophorus xiphidium.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorphi; Acanthopterygii; Percormorpha; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 NX NCBI_taxid=8086;
 RX MEDLINE=98241172; PubMed=9582016;
 RA DinitzIjevic N., Winkler C., Welbrock C., Gomez A., Duschl J.,
 RA Altschimed U., Scharf M.,
 RT "Activation of the Xmrk proto-oncogene of Xiphophorus by
 RT overexpression and mutational alterations.";
 RL Oncogene 16:1681-1690(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIO PURIFICATION;
 RA Scharf M.,
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U53471; AAD10500.2; .
 DR HSSP; P13362; 1PFG.
 DR InterPro; IPR000345; CytC heme bind.
 DR InterPro; IPR000494; EGFR L domain.
 DR InterPro; IPR000719; Euk_Pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR001245; Tyr_Pkinase.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF01030; Reced L domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR Prodom; PD000001; Euk_Pkinase; 1.
 DR SMART; SM00261; Fv; 3.
 DR SMART; SM00219; TyKc; 1.
 DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.
 DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 2.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.

SQ SEQUENCE 1165 AA; 129614 MM; 7F7EE38D8771A74E CRC64;
 Query Match 40.3%; Score 2730; DB 13; Length 1165;
 Best Local Similarity 45.8%; Pred. No. 2e-197;
 Matches 582; Conservative 163; Mismatches 383; Indels 142; Gaps 30;

```

  1 MELALCRWGLLALDPG-AAST---QVC---TOYIKANSKFTIGITELLVGGCGVQV 51
  4 LELLEL-----LLILLLLISIGRCSTDPDRKVCQGTSMQMTMLDHHYLMKMK-MYSGCNVLL 58
  52 GNLELYLPTNASLSFLODIQEVGYVLIHNOVROPIQRIQIVGCTQLEFDNVALATL 111
  59 ENLEITTTQENODLSFLOSIQEVGYVLIHMAVESTPLVNLLELRLGONVLEGENFLLWM 118
  112 DNGDPLNNTTPTVGASDGLRELQSLLEILKGVLIQNPOLCYQDTTILMKDIFHKAN 171
  119 SNYCK-NPSSP--DVYGVGKQQLSLNLTLEILSGYKVASHPPLCNVETINMDIVDKTS 175
  172 QALTLTIDNRSRACHCSPMKGSRCKWSSSDCQSLRTVYAGGC-ARCKGPLPTDCC 230
  176 NPTNMLIPLAFERQCCQKCPGCVNGSCWAPGPGHCQKFTKLCAEQCNRCRGPRIDCC 235
  231 HEQCAAGCTGPKHSDCLACLFHNSGICEILHCPALVTYNTDFESMPNDEGRYTFGACV 290
  236 NEHCAGGCTGPRATDCLACRDFNDGCTCPTPKIYDIHQVYVDNINIKTTGACV 295
  291 TACPYNVLTDSVGSCTLVCPJHNOEYTAEDGTORCEKSKPCARVCYGLMEHLREAV 350
  296 KECPSNVYVTE-GACVASCAGMLLEV-D-ENGKRSCKPCGCVGPKYCDGIGISLNTIAV 353
  351 TSANIOEFAAGCKKIFGSLAFLPESPDGDPASNTAPLOPQLOVFETLEITLYLTISMP 410
  354 NSTTIGSFNCTKINGDIILNRSFSDDPYKIGPDPHMLNLTIVKLTIVLYIMMP 413
  411 DSLPDLISVFQNLQVIRGLIHNGAYS-LTLQIGISWLGRLSRLREGSLALIHNTILC 469
  414 ENMTSLSFQNLLEIRGRITFSRGSFVVYVQVSHLQWLGLSLKEVSAQNVILKXTPQLR 473
  470 FVHTVPMDOLEFRPHQALHTANRPEDECVBELAHQICAGHCHGPRPTQCVNCSQFL 529
  474 YASTINMRRLFRSEDSIEYDAT-----EQCTNNKSSDGCWGPPTMCVSLHVD 526
  530 RQGEVCEECRVLOGLPREYVNAHCLPCHPECCPONGSVTFCGPEADOCVACAHYKDPF 589
  527 RGRGVASCNLLQGEPRBAQVDRVCVQHQECLYVDLSLTCVGPBANCSCKAHQQDQ 586
  590 CVARCPGCVKPDLSYPIWKFPREBEACOPCPINCHSCVDLDKGCRAEGCASPLTSIV 649
  587 CIRCPHGMIGDDTL-IMKYADKMGQCCPCHQNCCTQSGSPGLSGCRGD-IVSHSLAV 644
  650 SAVVGLLVVVGAVFGLIKRQOKIRKYTRRLIQLQETELVEPLTPSGAMPNQAOML 709
  645 GLVSGLLITYIVALLIVLIRRRIR-RKRTIRLLQELVEPLTPSQADNCFELIL 703
  710 KETELARKVYLSGAGFGTYKGIWIPDGENVKIPVAIKYLRENTSPKANKEILDEAYVA 769
  704 KETEFKXDRYLSGAGFGTYKGIWIPDGENIRIPVAIKYLRATSPKNOEVLDEAYVA 763
  770 GVGSPYVSRLLGLCLSTVQVLTQLMPYGLLDHYEENGRIGSODLLNMCQIAKGSY 829
  764 SYDHPVYCRLLGLCLSAVQVLTQLMPYGLLDYRQHEERICGWLNLWCQIAKGNV 823
  830 LEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGVPIKMALESIL 889
  824 LEERHLVHRDLAARNVLKPNHVKITDFGLSKLLTADKEVQAGKGVPIKMALESIL 883
  890 RRRFTQSDVMSGVTVMELMTGAKPYGIPARELPDLLEGELEPPPICTIDVYIM 949
  884 QMTYTHQSDVMSGVTVMELMTGSKPYGIPAKIASVLENGELPPPICTIEVYII 943
  950 VKCMIDSECRPFRELVEFSRMAADPQRFVYIENEDLGASPLDSTFYSLLEDDMG 1009
  944 LKCMIDPSSRPFRELVEFSQMAADPFRYIVIG---NLPSPEDRLFLFRLLSSD-- 998
  
```

QY 1010 DLYDAEEYLVQQGFFCPCDPAPCAGGVHHRSSSTRSGGDLTLGLEPSEBEARSP 106
 DB 999 DLYDADLYLL-----RYKRIN-RQS-----EFC 102
 QY 1070 APSEGAGSDVFDGDLGAKAGLQSLPTHDPSPRLQRYSDPTV-PLPSETDGVAPLTS 112
 DB 1022 IPRVGH-----PYRENSIALRYISDPTQNALKXOLDG----- 105
 QY 1129 PQPEYVNPDPVRPQ-----SPRE-----GPLP-AARPAATLERAKTISPKNGV 117
 DB 1055 ---EYVNPQSESTSSRLSDIYNNVEDLTDGQGPVLSQGEAETNFSRREYVNTQNSI- 111
 QY 1176 KDVFAFGAVNEPEYLTLPQGAAPRPPAPSPAFPDNLVYWDQDPPEGAPSTFGTP 123
 DB 1111 --PLVSGSSMDPDDY---QAG-----YQAAE-----LPQTGALTNQMFLP 114
 QY 1236 TARNPEYLG 1245
 DB 1147 AARNLELYGL 1156

RESULT 5

Q9W6F6 PRELIMINARY; PRT; 1137 AA.

AC Q9W6F6 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Receptor tyrosine kinase (Fragment).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archaeoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=HINDRAIN;
 RX MEDLINE=9263203; PubMed=10328884;
 RA Dixon M., Lumsden A.;
 RT "Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in
 RL embryonic chick hindbrain".
 DR EMBL; AF121963; AAD31764.1; -.
 DR HSP; P11362; IFCX.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR000719; Euk_dkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001368; TNFR_c6.
 DR InterPro; IPR001245; Tyr_kinase.
 DR InterPro; IPR004019; YLP_motif.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; kinase; 1.
 DR Pfam; PF01030; Recep_L_domain; 1.
 DR Pfam; PF02757; YLP_2.
 DR PRINTS; PRO0109; TYRKINASE.
 DR ProDom; PD000001; Euk_dkinase; 1.
 DR SMART; SM00261; FU; 3.
 DR PROSITE; PS00219; TYRK; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
 DR PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.
 KM Kinase; Tyrosine-protein kinase.
 FT NON_TER 1
 FT 1
 SQ SEQUENCE 1137 AA; 127927 MM; 4D616436F87DC84F CRC64;

Query Match 40.3%; Score 2729.5; DB 13; Length 1137;
 Best Local Similarity 47.3%; Pred. No. 2.1e-197;
 Matches 539; Conservative 169; Mismatches 354; Indels 77; Gaps 24;
 QY 155 LCYQDTILMKDIFHKNNQALTLIDNRSRACHCSPKCKGRCWGSSSDCQSLRTFTVC 214

Db	3	LCFADTIIHMODIVRNPMASNFTLVPNTGSSGCGCHKSCIG-RWGMTEHHCQTLLTITVC	61
Qy	215	AGGC-ARCKGRLPTDCCHEQCAAGTGKASDCLACHFNHSGICEILCPALVTYNTDTF	273
Db	62	AEQCDRCYGPVSSDCHRECAAGSGKPDNDCAFNMFNDSGACVQCQOTFEVNTTF	121
Qy	274	ESMPNREGRYTFGASCVCAPYVNTLSTVSGCTVCPILHNOEVAEDGTORCEKSPCA	333
Db	122	QLEHNHNAKTYTGACVCKCFHNFV-VSSSCVACACSSKMEV-EEHGKCKRPTDIDCF	179
Qy	334	RVCYGLGMEHLEVRVAVTASANIQEFAGCKTIFGSLAFIPESFDGDPASNTAPLOPEGLQV	393
Db	180	KACDGIQGSIVSAQTVSSNIDKFNCTKINGNLIFLVGIHDDPHYTAIAINPEKNTI	239
Qy	394	FETLEITGYLYISAMPDLPDLSPFQVLYIRGRILHNGAVSLTQGLISWLGRLSR	453
Db	240	FQTVREITGYLQGMFENMTDFRVSUVTYIGRALYSGLSLILKQOQITSLQFQSLK	299
Qy	454	ELSGGLALIHNTLCEVHTVPMQDLFRNPHQALLHTANPREDECVGEGLAGHQLGARGH	513
Db	300	QISAGNIYITDNSNLGYHTVNMWLSLSTPSQKTVIHRNKAENCTADGMVCHLSSDQ	359
Qy	514	CMGPGPTCCVNSQPLRBOECVEECRVLOGLPREVYANRHLPGHPCPC-QNGSVTCG	572
Db	360	CMGPEPDCLSKRIRKRTICBSCNLIDGFEFFAVGSVCMECDPCCEKMEENMTTCY	419
Qy	573	PEADQCVACAHYKDPFCVACRPGVXEDLSYMEIMKFPDEGACPCPINCTHSCVDLD	632
Db	420	PGPHCTKCFHFKQGNPCVEKCPDGLQANSF-IFRYADEBDEHCHCHNCQCGRGA	477
Qy	633	DKGC-----PAEQRASPLTYSVAV-GILLVVVLGVYFGLLRQOKTRKYT	680
Db	478	SHDCIYPMTRQSTLPHAR-TPV--TAAAGTIGLFTIIVMGTLFAYVARRKSIK-KKBA	533
Qy	681	MRLLQETELVEPLTPSGAMPNQAOMRLKETELRKYKVGSGAFVYVYGVIMPDGENV	740
Db	534	LRRL-ETELVEPLTPSGTANQALILKTELKRYUGSAGFVYVYGVIMPDGENV	592
Qy	741	KIPVALKVLRENTSPKANKELIDEAIVMAGVGSFVYVRLIGICLTSTVQLVQLMPYGL	800
Db	593	KIPVALKILNTTPKANVEFMDEALIMASMDHPLVRLGVCISPLQIVLQJLMPHGCL	652
Qy	801	LDHYRENRGLSGODLLNMCMQIAKMSYLEDYRLVHRDLAANVLYKSNHVKITDEGL	860
Db	653	LDYVHEKDNIGSULLNMCVQIKKMMYLEBRRLVHRDLAANVLYKSNHVKITDEGL	712
Qy	861	ARLIDIDETEVHADGKVPKIMWALSILRRRFTHSDVMSYGVTWELMTFGAKPYDGI	920
Db	713	ARLLEGEKEYNAGGKMPKIMWALECIIHYRKFTHQSDVMSYGVTWELMTFGAKPYDGI	772
Qy	921	PARLIDLEKGEPLPOPICTIDVYIMIKCMWIDSECPREFRELVSERMAARDPQRF	980
Db	773	PTREIPDLLEKGEPLPOPICTIDVYIMIKCMWIDSDRPFKELEAAESRMAARDPQRF	832
Qy	981	VVIONED-LGPASPLDSTFYRSLLEDDMGDLVDAEEYLVPOGFCFPPAPAGAGMWH	1039
Db	833	LVIQGDDRMKLPSPNDKFPQNLLEDEDLDDMMDAEEYLV-QAFNIPPIYTSKRIDS	891
Qy	1040	RHSSSTRSGGDLTLLEBSEEAAPS--PLAP-SEGASDVFDQDLGMAKGLQISLP	1096
Db	892	NRNOFYVRDGGYAAEGV-PMFYRAPGCIIPFAPVAGATAELFEDTCNGTLRKQVATL	950
Qy	1097	THDPSLQRYSEDPVLPSPS-----ETDGYVAPLTCSPQPEYVNPQVPRQPPSPRG	1149
Db	951	AKEDSSTQRYSDPTVIFRERVIKGLDEDEGVTMPKDKRKIDYLNPEVNEFVSRKNG	1010
Qy	1150	PLPAA-RPAGATLERAKTISPCKKGVKGVF-----AFGAVNPEYLTPOGGAAPQ	1200
Db	1011	DLQAVNDPEYHN-----APNGQRAEDEVYVNEPLYNTFANTLENAEYL-----K	1055
Qy	1201	PHPPAFSPAFDNLVYWDQDPPERGA--PSTFGKGT-----AENPEYL 1243	
Db	1056	NNLPKAKKAFDNDYNNHSLPPESTLQHEDLYOEYSTKFKYKQNGRIRPIVAENPEYL	1114
Qy	RESULT 6		
Db	P79754	PRELIMINARY; PRT; 1328 AA.	
AC	P79754		
DT	01-MAY-1997 (TRENBLREL.03, Created)		
DT	01-MAY-1997 (TRENBLREL.03, Last sequence update)		
DT	01-JUN-2002 (TRENBLREL.21, Last annotation update)		
DE	Etdb3.		
OS	Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;		
OC	Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;		
OC	Tetraodontidae; Takifugu.		
OK	NCBI_TaxId=31033;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=99177347; PubMed=10077531;		
RA	Geliner K., Brenner S.;		
RT	"Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu		
RT	rubripes".;		
RL	Genome Res. 9:251-258(1999).		
DR	EMBL; AF056116; AAC34391.1; -		
DR	HSSP; P11362; IEGK.		
DR	InterPro; IPR000494; EGFR_L domain.		
DR	InterPro; IPR000719; Euk_Dkinase.		
DR	InterPro; IPR001245; Tyr_Dkinase.		
DR	Pfam; PF00757; Furin-like; 1.		
DR	Pfam; PF00069; Dkinase; 1.		
DR	Pfam; PF01030; Recep_L domain; 2.		
DR	Prodom; PD000001; Euk_Dkinase; 1.		
DR	SMART; SMO0261; Fu_3.		
DR	SMART; SMO0219; Tyrc; 1.		
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.		
DR	PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.		
KW	ATP-binding; Transferase.		
SEQUENCE	1328 AA; 148613 MW; A333039258B647E9 CMC64;		
Query Match	33.8%; Score 2292; DB 13; Length 1328;		
Best Local Similarity	41.28; Pctd. No. 3.3e-164;		
Matches	510; Conservative 148; Mismatches 397; Indels 184; Gaps 29		
Qy	44	YQGQVQGNLELTYLTFTNASLSFLQDICEVQGYVLIANQVRQVPLQRIYRGTQLE	103
Db	51	YKGEIIMGNLEITQIESNMPDFLKTIREVTGYVLIAMNHPEIPLGQRYRGNSLTE	110
Qy	104	DNYLAVLDNGDPLNNTTPVTGASPGGLREQLRSUTLEILKGGVLIQRNPOLCYODTILM	163
Db	111	RRFALSVFLN-----YPKDG--PSGLNQLGIMNLTLELDGQVLIINKKYLRGPMVYM	161
Qy	164	KDIFKKNQALTLIDITNRSRACHPSPCKMGKSGRSGESSEDDQSILTRIVCAQGC-ARCK	222
Db	162	RDIL-RKNDAPRIETQPNRGENGVCH---KSC-GNYCWPGRKQDQILTKVCAPOCNDRCF	216
Qy	223	GPLEPTDCHEQCAAGCTGPRHSDCLACHFNHSGICELCPALVTYNTDTFESMPNREG	282
Db	217	GTSRSDCHIECAAGGKPLDTCFACRLFNDSGACVPCQPTLIYNKQTFQMEINPNMX	276
Qy	283	YTFGASCVTACPNVYLTVDGSGTLCVPLHNGEYVAEDGTOR-CEKCSFPCARVYCGGLM	341
Db	277	YQGISIVSQCPTHFV-VDSGSCVSVCPDKMEV--ERSQRCQCELCGSLCPKVEGTGA	333
Qy	342	EHRVAVAVTSANIOEFAGCKTIFGSLAFIPESFDGDPASNTAPLOPEGLQVFTLEBIT	401
Db	334	E--QRTGVSSNIDSFINCTKIQSLHFLVGIHDDPHYTAIAINPEKNTI	390
Qy	402	GVLISAMPSPDLSPDLSPFQVLYIRGRILHNGAVSLTQGLISWLGRLSGRLSGAL	461
Db	391	DIINTQSWPEKLENDLSVFSLLTIIQGRSLFKRFSIMVMKIPTLTSLGLRSLREISDGSVY	450

```

QY 462 IHANTHLCFTVTVEMDQLFRNPH-QALLHTANRPEDECVBGLACHOLCARGHCKWGPPT 520
Db 451 ISONALHCYHHTVMTOLFRGSRVRANSLSNRMACVADGVCDFPLCSDSCMCFPGD 510
QY 521 QCVNCSQFLRGQCEVEGCRVLOGLFREYVANH-CLTCHPECCOPONGSVYCFGPADOCY 579
Db 511 QCLSCRNSRKHGTCAVAGCHFNSSGIFREFAGANGVCAVACHECKRFGKASCTSPGADBCW 570
QY 580 ACAHYKDPFCVAPCPSGVKPDJLYMPIMKFPDEEGACOPCPINCTHSQVLDLDDKCCPAB 639
Db 571 ACTKFRDGPYCMSSCPAGVN-DGEKGLIFKFPNNEGCEPCHQCTGGCGPGLNDC--- 626
QY 640 QASPLTIVSAVVGILLVYLVGVF-----GILIKRQOKIRKXTMRLLQELTVE 692
Db 627 LEAALTISSQGITGIALGVAGLIFCLVIFLGLMLVHRGLAIRRKAMRYLESSESE 686
QY 693 PLTSGAMPNOAMRIKELTELKRYKVLGSGAFGTVYKGIWIPDGENVKIPVAKTLREN 752
Db 687 PLGP-GEKGTAVHARILKPSDLRKIKPLGSGVGTGSKGFWIPEGTVKIPVAKTIQDS 745
QY 753 TSPKANKEILDEAVVAGVSPYVSRLLGICLTSTVOLVQLMPYGLDHYENRGRIG 812
Db 746 SGRQTFEITDHLISMSGLHPYIVRLGICPGTCLQVLVQLSSHSLLEHTIRQHTSLD 805
QY 813 SQDLNMCQIAKXMSYLEVRLVHRDLAARNVLYKSPNFKITDGLARLIDIDETEH 872
Db 806 PQRLLNMCVOIAKMYLYLEHRVVKTLARNILKNDYQVQSLSDYGVADLIPDDKXIV 865
QY 873 ADGKVPKIMMALESILRRFTHQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKG 932
Db 866 YSEKTPKIMMALESILFRRYTHQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKG 925
QY 933 ERLPQPICTIDVYMIWYKCMIDSEGRPRELVSFESMAADPQRFVYIQWEDIGPAS 992
Db 926 ERLQOPALCTIDVYMWKCMIDENIRPTFKELASFTMAADPPRYLVIRNEG----- 980
QY 993 PLDSTFYRSLLEDDMDKDLVDAEZYLVQOGFCPPAPGAGGMHRRHSSTRSGGD 1052
Db 981 -----EDSGMGEFL-----RRGSR-----GL 997
QY 1053 LTLGLEPSEEFAPRPLAPSEGASDVFDDGLMG---AAKGILQSLPTHDPFLQ----- 1104
Db 998 LEADLEDEDEE-----GLGDRFATPSSLQPSFWSSTSPQINSYM 1036
QY 1105 -----RYSEDEPTVLPSETDGYVAFLTCSPP- EYVNO-----PDVAPCP 1144
Db 1037 VMTQLRYD-----FAVSGGHIGYLPMSPPVDITRQLMYQSRSLSVRTLPDRSAFR 1090
QY 1145 SPREGPL--PAAPAGATLEBAKTLSPGKNGVVDVAFGGAIVENPEYLLPQGAAPOPH 1202
Db 1091 SSRRAELCEDGAGCAGIFRYR-----FGSERGN-----POGG----- 1122
QY 1203 PPAPSPAFDNLVYWDQPPERGAPESTFKGTPTAENPE 1241
Db 1123 -----QQRKUSTASSPSSFTWMADEDE 1146

```

RESULT 7
Q9BIH9
AC Q9BIH9: PRELIMINARY; PRT: 1433 AA.
DT 01-JUN-2001 (Tremblrel, 17, Created)
DT 01-JUN-2001 (Tremblrel, 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel, 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:
OC Pterygota: Neoptera: Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anophelinae
CX NCBI_taxid=7165;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN-SUA;
RA Lyceat G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301655; CACJ5008.1; -.
DR HSBP; F11362; IFEK.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PS00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU_7.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00103; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON_TER
SQ
SEQUENCE 1433 AA; 159585 MW; E3D9D8896724F07 CRC64;

Query Match 29.5%; Score 2000.5; DB 5; Length 1433;
Best Local Similarity 32.6%; Pred. No. 4.2e-142;
Matches 464; Conservative 191; Mismatches 383; Indels 385; Gaps 36

QY 44 YQCGVAVQVNSLELTYPETNASLFLDIOGVQVLIHANQVQVRLQRLRYGTQLF- 102
Db 25 YTNCTYDGNLEITRINIQTDLNPLAHIREVTGYVLSLYDLQVILLPRQITRGHTFK 84
QY 103 -----EDNYALAVLDNDPPLNNTTPVVGASPGRLQLSLTLLIGVYLICNPOLCVO 158
Db 85 LNKWEAYGLFV-----SSSHNTLELPRLRITLGSVGFNNVNLCH 128
QY 159 DTLIMKDI-FKKNQALTLIDITNSRACHCSPMKGSRCESSSEDCSLTRTVACAG 217
Db 129 KSIWEEIILAPQSMQYTFNFSRPERVCPCHPSCSEVG-CWEGAHNQRPSKLNCSQ 187
QY 218 CA--RCKGRLPTDCHEQCAAGCTGPKHSQCLALHPNHSIGICELHPALVTYNTDFES 275
Db 188 CSQGRCPGPKPRECHCFLFCAGGCTGFLQSDCLACKNFYDGVCKQCECPMQIYNPTNFW 247
QY 276 MPNPEGRYTFGASGVTAAPYNYISTVSGCTVCPILNQEVTAEDGTQCEKSKCAEV 335
Db 248 EPNPDGKXAGATCYRKP-EHLIKNGACVRCSPGKKNPQNSE-----CVPEKGVCPXT 301
QY 336 CYGLGMEHLREVAVTSANIQEFAGCKKIFGSLAPLPESFPDGPASNT-----APLQ 387
Db 302 CPGEGLVH-----SDNIGNYKDCITIEGSLIEDQSFDFGQVYTFNFSGPRYIKTD 353
QY 388 PEOLQVETLEETITGYLYISAMDSLDLSVFNQLOVITGRILHNAV-SLTLOGIGIWM 446
Db 354 PDRLEVSTYKEITGFINTQAHHPNTTLYNFENLEVVGRQKENLFAVYIVKTSLS 413
QY 447 LGLSLRELGSGLALIHNTHLCFTVTVEMDQLFRNPHQALHTANRPEDECVBGLACH 506
Db 414 LEUKSLKRVNVSGLIVLENSDLDFVEDIMSEIKSSDHEWVQKNRNATTEHEEMES 473
QY 507 QLCARGHCMWPGTQCVNCSQFLRGQCEVEGCRVLOGLFREY-VNARNLPHQPCOPON 565
Db 474 EQGSKAGCWKGEGQCECKNRYKXKCLDSCK--SLPRLYVSVSKTGCDHOECKD-- 528
QY 566 GSYTCFPGPADQCAVAHYKDPFCVAPR----- 595
Db 529 ---FCYGNEDNCGSCMNVYDGFCAVABCPRTKMANNGTCINCHKTVCGRGPRDTIAPD 585

```

```

QY 596 -----SGVKPDLSTYMPIMKFPD-----BEG----- 615
DB 586 GCISCDKAIIGSDAKIERCLMKDESCDPGYSDVYLQEGGLKOLSGAVCRKCHPCKK 645
QY 616 -----ACQPCPINCT-----HSC 628
DB 646 CTGYGHEQFCQECTGYKKGEQCEDECEQDFYANEETRICTPCHQEGCGHGLDDHEC 705
QY 629 VDL-----DKCPAEQ-----RASPLTSI 648
DB 706 RLKXLFEGDPYNATFTTCVSNCPASHYKFPQAGKIGYGCADSNQSLREPTQV 765
QY 649 VSAVVGILLVVLGVVFGI---LIRROQKIRKYMRLLOTELVELPELTGSGAMPNQAQ 705
DB 766 KIVMGSVMALLILCVFGLIAFVLFGRHKKKDAVMKTMALAGCDESEPLRPSNVGPNLTK 825
QY 706 MRILKETELRKVYVGSAGFTVYKGIPIPGENKIPVATIKULRENTSPKANEILDEA 765
DB 826 LRIIKEAEIRGCVLQMGAFGRVFGVMVPEESVKIPVALKVLMEVSGSESSKEFLEEA 885
QY 766 YVMAGVSPYVSRLLIGICTSTVQVLTQLMPEYGCILDHVRNQRIGSODLLNMQOIAK 825
DB 886 YIMASVEHENLKLKLAVCMSTQMKLITQLMPLGLDLYVRNKKKIGSKALLNMQOIAK 945
QY 826 GMSYLEDVRLVRLDAARNVLYKSPNHKIDTFGLARLIDETERYADGKVPYIKMAL 885
DB 946 GMAVLEERLVARDLAARNVLYQTPSCVKITVFGIAKLDPDSDEYAAAGKMPKIMLAL 1005
QY 886 ESILRRRFTQSDVSYGVTVWELMTFGAKPYDGIIPAREIPDLKEKERTPOPICITDV 945
DB 1006 ECIRHVRFTSKSDVAFGITITWELTYGARPYENVPAKDVPELLEIGHKLPQDPCIGLDV 1065
QY 946 YMIWKCWMIDSECPREFRELVSERAPORFVYIQNEDELCPASPLDSTFYRSLIED 1005
DB 1066 YCILLSCWLDADAPTEKOLAETFAEKARDPGRYLM-----PGDKFMRLPSYTN 1116
QY 1006 DDMGDLV-----DAEYLVFQCGFCFCDPAPAGCM 1036
DB 1117 QDEKDLIRTLAPVMAAAAAAAAAAGASVNDVPSITAEIDELQKTPSITMLPEPSA--- 1173
QY 1037 VVHHRSSSTSGGGDLTLGLPESEEAPRS-----PLAP---SEGAGSDVFDGLGNG 1087
DB 1174 -----VEPS-DEMFKSLRYCKXPLKRPDETDGHEV-----GVC 1207
QY 1088 AAKGLQSLPTHDPSPLORESDEPTVPLPSEITDGVAVLTCSPQREYVNCQDVAPQPPSPR 1147
DB 1208 GIR-----LNLPLDEDDYLMPTCOSQ---NOS----- 1231
QY 1148 EGPLPAPAPAGATLERAKTISPGKNGVVKVFAFGAVENPEYL-----TPQGGAA 1198
DB 1232 -----TPG-----YMDLIGVPAVDNPEVLMGSTQALAGLAGSGMG 1267
QY 1199 PQPHPPAPFPAFNLVYWDPPERGAPESTFGTGTAEINPE 1241
DB 1268 --PHTP-----PPTVNGMPTHQHSQ 1287

RESULT 8
QYUK79 PRELIMINARY; PRT; 419 AA.
AC QYUK79:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;

```

```

RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).
RN 12
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AR177761; AAD56009.2;
DR InterPro; IPR00494; EGR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like.
DR Pfam; PF01030; Recept_L_domain; 1.
DR SMART; SM00261; Fu; 1.
SQ SEQUENCE 419 AA; 45472 MW; FECEB347E2D0030C CRC64;

```

```

Query Match 25.7%; Score 1740.5; DB 4; Length 419;
Best Local Similarity 84.3%; Pred. No. 3,26-123;
Matches 332; Conservative 9; Mismatches 40; Indels 13; Gaps

```

```

QY 1 METALGSMGLLALPPGASTQVCT-----QYKANSKFIGITELLYGCGVYQGNL 54
DB 1 METALGSMGLLALPPGASTQVCTGTDMKRLPPSPETHLDMRLHYGCGVYQGNL 60
QY 55 ELTYLPTNASLFIQDIOEVGYVLIANQYQVPLQRLRIVRSTQLFEDNYALAVLDNG 114
DB 61 ELTYLPTNASLFIQDIOEVGYVLIANQYQVPLQRLRIVRSTQLFEDNYALAVLDNG 120
QY 115 DPLNNTVTGASRGULRELOSLTELKRGVLIQNPOLCTQDTILMKDIFRKNOLA 174
DB 121 DPLNNTVTGASRGULRELOSLTELKRGVLIQNPOLCTQDTILMKDIFRKNOLA 180
QY 175 LTLIDTNSRACHPCSPMKCKSRGWCSSSEDCQSLRTVACAGCARCKGFLPTDCHEQC 234
DB 181 LTLIDTNSRACHPCSPMKCKSRGWCSSSEDCQSLRTVACAGCARCKGFLPTDCHEQC 240
QY 235 AAGCTGPHSGCLACLPNHSIGTELHPALVTYNTDFESSMNPBGRYTFGASCYTACP 234
DB 241 AAGCTGPHSGCLACLPNHSIGTELHPALVTYNTDFESSMNPBGRYTFGASCYTACP 300
QY 295 YNVLSTVGSCTLVCPHNOVTAEDGTQCEKSKRCARVCGLGMEHLREYRAVTSAN 354
DB 301 YNVLSTVGSCTLVCPHNOVTAEDGTQCEKSKRCARVCGLGMEHLREYRAVTSAN 355
QY 355 IQEPAGCKKIFGSLAPFESFDGDPASNTAPLP 388
DB 356 LRMQPG--PAHPVLSFLRPSGMDVSAFYSPLAP 387

```

```

RESULT 9
QYR2X1 PRELIMINARY; PRT; 367 AA.
AC QYR2X1:
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN 11
RP SEQUENCE FROM N.A.
RA Straube R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC027080; AAR27080.1; -.
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

```

```

Query Match 25.7%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 3,46-123;
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0.

```

```
QY      883  MALSLIRRRFTHSDVWSYGVYWEIMTGAKPYDGIIPAREIPDLLEKGERLPQPICT 942
Db      1  MALESILRRRTTHSDVWSYGVYWEIMTGAKPYDGIIPAREIPDLLEKGERLPQPICT 60
QY      943  IDVYIMVWKCMIDSECRPFRELVSFSRMARDPQRFVYIIONEDLGSPASPLDSTFYSL 1002
Db      61  IDVYIMVWKCMIDSECRPFRELVSFSRMARDPQRFVYIIONEDLGSPASPLDSTFYSL 120
QY      1003 LEDDDMGDLVDAEEYLVPOQGFPCPDPAFGAGMHHRRHSSSTRSGGDLTLGLEPSEE 1062
Db      121 LEDDDMGDLVDAEEYLVPOQGFPCPDPAFGAGMHHRRHSSSTRSGGDLTLGLEPSEE 180
QY      1063 EAPRSPLAPSEAGSDVDFDDLAVGVTKGLQSPLDLSPLQRYSDPTLPLPPTDGVY 1122
Db      181 EPPRSPLAPSEAGSDVDFDDLAVGVTKGLQSPLDLSPLQRYSDPTLPLPPTDGVY 240
QY      1123 APLTCSQPPEYVNOPDVRPOPSPREGPLPAARPAGATLERAKTLSPGKNGVYKDVFAFG 1182
Db      241 APLTCSQPPEYVNOPDVRPOPSPREGPLPAARPAGATLERAKTLSPGKNGVYKDVFAFG 300
QY      1183 GAVENPEYLPQGGAAQPPHPPAFSPAFDNLVYWDODPPERCAPSTKGTPTANPEY 1242
Db      301 GAVENPEYLPRAQTASQPHSPAFSPAFDNLVYWDONSECGPPSTTEGTPTANPEY 360
QY      1243 LGLDVPV 1249
Db      361 LGLDVPV 367

RESULT 10
086712  PRELIMINARY; PRT; 729 AA.
AC 086712
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polyprotein.
GN POLYPROTEIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
CX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; Pubmed=8152791;
RA Venustrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR HSSP; P03322; AAC60725.1; -.
DR HSSP; P03322; 1A6S.
DR InterPro; IPR000719; Euk pkinase.
DR InterPro; IPR004028; Retro M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase.1.
DR Pfam; PF02813; Retro_M_1.
DR ProDom; PD000001; Euk_pkinase.1.
DR SMART; SM00219; TyrKc.1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D2FF914EFD1D63 CRC64;
```

Query Match 25.4%; Score 1720; DB 15; Length 729;
Best Local Similarity 54.8%; Pred. No. 2,6e-121;
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;

```
QY      563  PONGSVTCFPEADQCAHYDPPFCVACPSGVKPDLSYMPIWKPFDEAGACQCPPI 622
Db      141 PEETATPKTGP--DHCKCAHFLIDGFHCYVACPAVLGENDTL-VWKYADANAVQCLHP 197
```

```
QY      623  NCHSCVLDLDDKCPAPQAPLSTSVSAVY-GILLVVLGVVFGILIRROCKIRKTYM 681
Db      198 NCRGCKCPGLECP---NSKTPSIAGVGGCLCLVYVGLIGLYLRR-HYAKRTL 253
QY      682  RRLQERLEVPILTPSGAMPNQNOMLLETETLRKTVYGSAGFVNVGIMTPDENVK 741
Db      254 RRLQERLEVPILTPSGAMPNQNOMLLETETLRKTVYGSAGFVNVGIMTPDENVK 313
QY      742  IPVAIKYLRNTSPKANKELIDEAVMAGVSPYRLLGI CLTSTVOLTOIMPYGCLL 801
Db      314 IPVAIKELRNTSPKANKELIDEAVMAGVSPYRLLGI CLTSTVOLTOIMPYGCLL 373
QY      802  DHRNENGRIGSQDLNMCQIKKMSYLEDVVLVARDLAARVLYKSNNHKKITDFGLA 861
Db      374 DYLRHNDNGISQYLLNMCVQIKGMVYEEERLVRDLAARVLYKTPQHVAKITDFGLA 433
QY      862  RLIDIDETVHAQGVKPIKMALESILRRFRTHSDVWSYGVYWEIMTGAKPYDGI 921
Db      434 KLGADSKETHAEGKVPFKMALESILRRFRTHSDVWSYGVYWEIMTGAKPYDGI 493
QY      922  AREIPDLLEKGERLPQPICTIDVYIMVWKCMIDSECRPFRELVSFSRMARDPQRFV 981
Db      494 AREISVLEKGERLPQPICTIDVYIMVWKCMIDSECRPFRELVSFSRMARDPQRFV 553
QY      982  VIQ-NEDLPASPLDSTFYRSLLEDMDMGDLVDAEEYLVPOQGFPCPDPAFGAGMHH 104
Db      554 VIQDERMHLPSPTDSKFRYRTLEEDMEDIVDADEYLVPHQGF----- 598
QY      1041 HRSSSTRSGGGLTLGLEPSEEPAPSPPL-----APSEAGSDVDFDDLGMAKAGLQSL 109
Db      599 -NSPT-----STRPLLSLSATSNNSATNCID-----RNGQGH 631
QY      1096 PTHDPSPLQRYSSDPTVPLPSET--DGVAAPLTCSQPPEYVNOPDVRPOPSPREGPLPA 115
Db      632 PVEDSIVQRYSSDPTGNGFLSESIDGFL-----PAPEYVNO--LMPKPS----- 675
QY      1154 ARPAGATLERAKTLSPGKNGVYKDV-----AFGAVENPEY 1191
Db      676 -----TAVNQNIYNNISLTALSKLPMDSRYQNSHSTAVDNPYL 715
```

```
RESULT 11
086714  PRELIMINARY; PRT; 567 AA.
AC 086714
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
CX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; Pubmed=8152791;
RA Venustrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR HSSP; S69372; AAC60727.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase.1.
DR Pfam; PF00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase.1.
DR SMART; SM00219; TyrKc.1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
```


KW Tyrosine-protein kinase.
 FT NON TER. 1
 SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7AD725E1 CRC64;
 Query Match 25.4%; Score 1718; DB 15; Length 567;
 Best Local Similarity 55.4%; Pred. No. 2,5e-121;
 Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 572 GPEADQVACAHKYPFCVACRCSGVAPDLSYPMKFPDEGACPCPINTCHSCVDL 631
 DB 1 GP--DHCKCAFLIGHPCVACRPAVGENDTL-VMKYADANAVCCLCPNCTRGKGP 57
 QY 632 DDXGCPAECRASPLTSISAVV-GILVVVLGVVPGILIKRQCKIRKYTWRRLLQETEL 690
 DB 58 GLEGCP--NGSKTPSIAGVVGGLCLVVGGLGILYLR--HIVKKTLLRLLQEREL 113
 QY 691 VEPLTPSGAMPVQAMRIKETELRKVVLGSGAGFYGYKIMVPDENAKIPVAKVLR 750
 DB 114 VEPLTPSGEAPVQALRIKETELRKVVLGSGAGFYGYKIMVPEGEKXKIPVAKVLR 173
 QY 751 ENTSPKANEILDEAYVAVAGVSPVSSHLCITSTVQLVTLQMPYGLLDHVRNGR 810
 DB 174 EATSPKANEILDEAYVAVAGVSPVSSHLCITSTVQLVTLQMPYGLLDHVRNGR 233
 QY 811 LGSODLMMKQIAKNGSYLEDVRLVHRDLAARVVLKSPHVKITDPEGLARLLDDETE 870
 DB 234 IGSQYLLMVCQIAKNGSYLEDVRLVHRDLAARVVLKSPHVKITDPEGLARLLDDETE 293
 QY 871 YHADGKVPKIMMALESILRRRPTHOSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLE 930
 DB 294 YHAEKGKVPKIMMALESILRRRPTHOSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLE 353
 QY 931 KGERLPPICTIDVYMVMKCMIDSECRPRELVSFSSMAADPQRFVYVQ--NEDLG 989
 DB 354 KGERLPPICTIDVYMVMKCMIDSECRPRELVSFSSMAADPQRFVYVQ--NEDLG 413
 QY 990 PASPLDSTFYRLLEDMDGLVDAEYLVPQGFPCPDPAAGAGVHRRHSSSTRSG 1049
 DB 414 LPSFTDSKFYRLTMEEDMEDIVDAEYLVPQGF-----NSPST--- 454
 QY 1050 GGDLLTGLPESEEPSPPL-----APSEGAGSVFPGDGLGMAKGLQSLPTHSDPQL 1104
 DB 455 -----SRTPFLSSLSATSNNSATNCID-----RNGQGHPRVEDSFVQ 491
 QY 1105 RYSEDPVPLPSET--DSYVAFLTCSPOPEYVNPDPVPPSPRPREPLPAARPAATLE 1162
 DB 492 RYSSDPVPLPSET--DSYVAFLTCSPOPEYVNPDPVPPSPRPREPLPAARPAATLE 526
 QY 1163 RAKTISPGRNGVVKVDF-----AFGAVENPEYL 1191
 DB 527 -----TAWVQNIYNNISITAIKSLPMDRYQNSHSTAVDNPVYL 566

RESULT 12
 Q8WYV0 PRELIMINARY; PRT; 412 AA.
 AC Q8WYV0;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Hypothetical 44.7 kDa protein.
 GN PF3659;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
 RA Man D.F., Gu J.R.;
 RT "Novel human cDNA clones with function of inhibiting cancer cell
 RT growth";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF318349; AAL5856.1; -
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR000719; Euk_kinase.
 DR InterPro; IPR001245; Tyr_kinase.
 DR InterPro; IPR004019; YLP_motif.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF02757; YLP; 2.
 DR ProDom; PD000001; Euk_Pkinase; 1.
 DR SMART; SM00219; TyrKc; 1.
 DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;
 Query Match 25.1%; Score 1697.5; DB 4; Length 412;
 Best Local Similarity 80.5%; Pred. No. 5,5e-120;
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4

QY 883 MALSIILRRRPTHOSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPICT 942
 DB 1 MALSIILRRRPTHOSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPICT 60
 QY 943 IDVYIMVMKCMIDSECRPRELVSFSSMAADPQRFVYVQ--NEDLG 100
 DB 61 IDVYIMVMKCMIDSECRPRELVSFSSMAADPQRFVYVQ--NEDLG 120
 QY 1003 LEDDDMGDLVDAEYLVPQGFPCPDPAAGAGVHRRHSSSTRSGGDLTLGLEPSEE 106
 DB 121 LEDDDMGDLVDAEYLVPQGFPCPDPAAGAGVHRRHSSSTRSGGDLTLGLEPSEE 180
 QY 1063 EAPSPPLAPSGASDVFDGDLGMAAKGLQSLPTHDPSTLQYSSDPVPLPSETDGYV 112
 DB 181 EAPSPPLAPSGASDVFDGDLGMAAKGLQSLPTHDPSTLQYSSDPVPLPSETDGYV 240
 QY 1123 APLTCSPOPEYVNPDPVPPSPRPREPLPAARPAATLEAKTILSPGRNGVVKVDFAFAG 118
 DB 241 APLTCSPOPEYVNPDPVPPSPRPREPLPAARPAATLEAKTILSPGRNGVVKVDFAFAG 300
 QY 1183 GAVENPEYLTPQGAAPQP-----HPPPA---FSPAFLNL 121
 DB 301 GAVENPEYLTPQGAAPQP-----HPPPA---FSPAFLNL 360
 QY 1215 YVMD-QDPPER-----GAPSTFGGTAA 1239
 DB 361 YVMD-QDPPER-----GAPSTFGGTAA 410

RESULT 13
 Q64895 PRELIMINARY; PRT; 962 AA.
 AC Q64895;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Gag-V-erb-A-V-erb-B protein.
 GN GAG-V-ERB-A-V-ERB-B.
 OS Avian erythroblastosis virus.
 OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
 OX NCBI_TaxID=11861;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=90206603; Pubmed=1969616;
 RA Bruskini A., Jackson J., Bishop J.M., McCarty D.J., Schatman R.C.;
 RT "Six amino acids from the retroviral gene gag greatly enhance the
 RT transforming potential of the oncogene v-erb-B";
 RL Oncogene 5:15-24(1990).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 DR EMBL; X52209; CA36459.1; -
 DR EMBL; X52211; CA36459.1; JOINED.
 DR HSSP; P10828; NTL.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR000536; Hormone_rec_119.

DR InterPro: IPR001723; StcHrmn_receptor.
 DR InterPro: IPR001245; Tyr_kinase.
 DR InterPro: IPR001628; Znf_C4steroid.
 DR Pfam: PF00104; hormone_rec. 1.
 DR Pfam: PF00069; kinase. 1.
 DR Pfam: PF00105; Zf-C4. 1.
 DR PRINTS: PR00398; STRDHOMNER.
 DR PRINTS: PR00047; STROIDFINGER.
 DR ProDom: PD000001; Euk_kinase. 1.
 DR ProDom: PD000035; Znf_C4steroid. 1.
 DR SMART: SM00430; HOL1. 1.
 DR SMART: SM00219; TykC. 1.
 DR SMART: SM00399; Znf_C4. 1.
 DR PROSITE: PS00031; NUCLEAR_RECEPTOR. 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP. 1.
 DR PROSITE: PS00011; PROTEIN_KINASE_DOM. 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR. 1.
 DR ATP-binding; DNA-binding; Nuclear protein; Receptor;
 KW Transcription regulation; Transferase; Tyrosine-protein kinase;
 KW Zinc-finger.
 SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E4E95CE CRC64;

Query Match 24.4%; Score 1653.5; DB 15; Length 962;
 Best Local Similarity 51.7%; Pred. No. 4.2e-116;

Matches 358; Conservative 73; Mismatches 142; Indels 119; Gaps 18;

QY 535 VEECRVLQGLPRE-YVNAR-HCLP-----CHPECO 562
 DB 354 IEKQSEYLLAFEHYINRKINIPHFWSKILMKVADLRMIQAVHASFHLMKVECTELS 413
 QY 563 PONGSVTCFGEAQCACAHYKOPPCVAPCPGSGVPDLSTYPIKFPPEBACQPCPI 622
 DB 414 PGE-----VGP--DHCKMCAHFIDGPHCVACPGVIGENDTL-VWKYADANAVCOLCHP 465
 QY 623 NCHSCVDLDKQCPAQRASPLTISVSAV-GILLVVLGVVFGILLIRROOKIRKRYM 681
 DB 466 NCTRGCKGPGLEGCP---NSKTPSIAAGVVGGLCLVVGGLGIGLYLRR-HIVKRTL 521
 QY 682 RLLIQTETLVEPLPPSGAMPNQAQMLIKETELRKVKVLCGSAFGVYVYKGIWIDENYK 741
 DB 522 RLLQERELVEPLPPSGAPNQAHLRIKTEPFKRVVLFQGAFGVYVYKGLMIPBEKXT 581
 QY 742 IPVAIKVLRENTSPRANKETIDEAYVAVGSPYVSRLIGICTSTVOLTQMLPFGCLL 801
 DB 582 IPVAIKELRENTSPRANKETIDEAYVAVSDNPHVCRLLGICLTSTVOLTQMLPFGCLL 641
 QY 802 DHVRENRGLSGQDLNWCQIAKGSYLEDVRLVREDLAAENVLYKSPNHVKITDFGLA 861
 DB 642 DYIREHKNIGSQYLNMVQVQIAKGNVLEERHVRHDLAARNVLYKTPQHVXITDFGLA 701
 QY 862 RLIDIDEYHADGGKVPKIMMALESILRRFTHOSQVMSYGVTVWELMTFGAKPYDGI 921
 DB 702 KOLGADEYHABGKVPKIMMALESILHRIYTHOSQVMSYGVTVWELMTFGSKPYDGI 761
 QY 922 AREIPDLLEKGERLPQPICTIDVYIMVWCMMIDSECRPRELSEFSRMAPRQFV 981
 DB 762 ASEISSVLEKGERLPQPICTIDVYIMVWCMMISGADSRKPFRELLAEFSKARDPRL 821
 QY 982 VIO-NEDLGPASPLDSTFYRSLLEDMDMGLVDAEYLVPQGFPCPDPAAGAGVTHR 1040
 DB 822 VIOGDERMHLPSPTDSFRYRLMEEDMEDIVDAEYLVPHQGF----- 866
 QY 1041 HRSSSTRSGGDLTLGLPESEEAAPRSLAPSEAGADVDFDGLGMAKGLQSLPTHP 1100
 DB 867 -NSPST-----SRPLSLSLATSN-----NSAKTKIDRNGH-- 898
 QY 1101 SPLORYSEDPVLPSTDTGVVALTSSPPPEYVNOQDVAPQSPREBPPLPAAPAGAT 1160
 DB 899 -----PVREDFL-----PAPEYVNO--LMPKPSSTAMVQNOIYVYISLT 936
 QY 1161 -LERAKTLSPKNGVAVKDVAFGAVENPEYL 1191
 DB 937 AISKLPWDSRYQN-----SHSTAVDNPEYL 961

RESULT 14

ID 085468 PRELIMINARY; PRT; 545 AA.

AC 085468; (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Avian Erythroblastosis virus (T834) v-erbB gene.

OC Avian erythroblastosis virus.

OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.

OX NCBI_TaxID=11861;

RN [1]

RP SEQUENCE FROM N. A.

RX MEDLINE=88217326; PubMed=2697102;

RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;

RT "Common site of mutation in the erbB gene of avian erythroblastosis

RT virus mutants that are temperature sensitive for transformation";

RL Oncogene Res. 1:265-278(1987).

DR EMBL; X06943; CA330024.1; -

DR HSSP; P11362; ITCX

DR InterPro: IPR000719; Euk_kinase.

DR InterPro: IPR001245; Tyr_kinase.

DR Pfam: PF00069; kinase. 1.

DR ProDom: PD000001; Euk_kinase. 1.

DR SMART: SM00219; TykC. 1.

DR PROSITE: PS00107; PROTEIN_KINASE_ATP. 1.

DR PROSITE: PS00011; PROTEIN_KINASE_DOM. 1.

DR PROSITE: PS00109; PROTEIN_KINASE_TYR. 1.

KW ATP-binding; Transferase; Tyrosine-protein kinase.

KW SEQUENCE 545 AA; 60899 MW; 140DCE8C6A0F8AF4 CRC64;

Query Match 24.3%; Score 1645; DB 15; Length 545;
 Best Local Similarity 54.9%; Pred. No. 7.8e-116;

Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15

QY 572 GPEADQCVACAHYKOPPCVAPCPGSGVPDLSTYPIKFPPEBACQPCPINTHSCVDL 631
 DB 1 GP--DHCKMCAHFIDGPHCVACPGVIGENDTL-VWKYADANAVCOLCHPNTROCKGP 57
 QY 632 DDKQCPAQRASPLTISVSAV-GILLVVLGVVFGILLIRROOKIRKRYMRLLOETEL 690
 DB 58 GLESCP---NSKTPSIAAGVVGGLCLVVGGLGIGLYLRR-HIVKRTLRELLQEREL 113
 QY 691 VEPLTPSGAMPNQAQMLIKETELRKVKVLCGSAFGVYVYKGIWIDENYKIPVAIKVL 750
 DB 114 VEPLTPSGAMPNQAQMLIKETELRKVKVLCGSAFGVYVYKGIWIDENYKIPVAIKEL 173
 QY 751 ENTSPKANKETIDEAYVAVGSPYVSRLIGICTSTVOLTQMLPFGCLLDHVRENRGR 810
 DB 174 EATSPKANKETIDEAYVAVSDNPHVCRLLGICLTSTVOLTQMLPFGCLLDVIREHKN 233
 QY 811 IGSQDLNMCQIAKGSYLEDVRLVREDLAAENVLYKSPNHVKITDFGLARLIDETE 870
 DB 234 IGSQYLNMVQVQIAKGNVLEERHVRHDLAARNVLYKTPQHVXITDFGLADEXE 293
 QY 871 YHADGKVPKIMMALESILRRFTHOSQVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 930
 DB 294 YHABGKVPKIMMALESILHRIYTHOSQVMSYGVTVWELMTFGSKPYDGIIPASEISSVLE 353
 QY 931 KGERLPQPICTIDVYIMVWCMMIDSECRPRELSEFSRMAPRQFVYIO-NEDLG 989
 DB 354 KGERLPQPICTIDVYIMVWCMMISGADSRKPFRELLAEFSKARDPRLVYIOGDERM 413
 QY 990 PASPLDSTFYRSLLEDMDMGLVDAEYLVPQGFPCPDPAAGAGVTHRRARRSSSTRSG 104
 DB 414 LPSPTDSKPYRLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
 QY 1050 GGDULTLGLPSEEAAPRSL-----APSEAGADVDFDGLGMAKGLQSLPTHPSPLO 110
 DB 455 -----SRPLSLSLATSN-----H----- 481

QY 1105 RYSEPTVLPSEETDGYVAPLTCSPQPEYVQNPDPSPSREGPLPAAPAGAT-LEP 1163
DB 482 -----PVRGEGFL-----PAPRYVQ--LMPKKEPSTAVQVQIYINYSITLAIK 523
QY 1164 AKTSPGKGVVQVAFAGCAVENPEYL 1191
DB 524 LPMDSRYON-----SHSTAVDNPEYL 544

RESULT 15
Q9WVF5 PRELIMINARY; PRT: 655 AA.
AC Q9WVF5;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., Maizle N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTAJ;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J., Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D., Maizle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egfr transcripts encoding truncated receptor isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai U., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishi Y., Arikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I., Saito T., Okazaki Y., Gojopori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuell P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., Schiml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Bazh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombauts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Wenz C., Whitaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S., Hayashizaki Y.;
RA "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001);
DR EMBL; AF124513; AAD44149.1; -
DR EMBL; AF275366; AAG28047.1; -
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AK004944; BAB23668.1; -
DR EMBL; AK004883; BAB23641.1; -

DR EMBL; AK004911; BAB23662.1; -
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR SMART; SMO0261; Fu; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 22.2%; Score 1503.5; DB 11; Length 655;
Best Local Similarity 43.9%; Pred. No. 5,2e-105;
Matches 281; Conservative 99; Mismatches 229; Indels 31; Gaps 9

QY 11 LVLALPAGAA--STGYC-----TGYIANKKFGITELLYQGQGVQGNELTYLPT 61
DB 14 LITLCAAGALEKKKQCGTSNLTQLGFEDHFSL-ORMTNCEVVLGNLEITYVOR 72
QY 62 NASLFLQDIOEVQGVYLIHNOYRQVPLRLRIVRGTOULFEDNALAVLDNGDPLNNTT 121
DB 73 NYDSLFLKTIQEVAGVYLIALTVERIPLENLQIRGNALYENTVALALISN----- 124
QY 122 PYTGASREGIARELOBSLTTELKGVYLQKNPOLCYODITLMKDI----FKKNQLALTL 177
DB 125 --YGNRTGLRELPKRNLOELILGAVRFSNNPILCNMDITQWRDIVQNVFMSMSMDL-- 180
QY 178 IDTRSPACHPSCSPMGKSGRSGSSSEDCSLTRTYCAGGCA-RCKGPLPTDCRQCAA 236
DB 181 --QGHPSKPCKDCSCGNSGCGWGGEGNCQKIKIICQAQCSIRCKRGSRSDCHQCAA 238
QY 237 GCTGPKNSDCLALCFNHSIGICELHCPALVTYTTDFESMPNPEGRYTGASCVTAQPN 296
DB 239 GCTGPRSDCLVCCGFODEATCKDTCPMLVNPPTYQMDVNEGKYSFGATCVKCKSPEN 298
QY 297 YLSTDVSGCTLVCPLIHQEYTAEDGTORCEKSPCARVYVGMENHLEVRVATANTQ 356
DB 299 YVTDHSGCVRACGPDYEV-EDDGIRKCKCKGPPRKVCNGIGIEFKOTYSINATNIK 357
QY 357 EFAGCKKIFGSLAFIPSSFDGDPASNTAPLPBPOLQVFTLEETITLYLISAWPDSLPL 416
DB 358 HFKYCTAISGLHILPVAFKGDSFTTRPLDPRELEITLKVKETITFLLIQAMPDWTDL 417
QY 417 SVFQNLQVITIGRIILHNAVSLTLOGLISMLGLRSRLBELSGSLALHNHTLCLFVHTVM 476
DB 418 HAFENLEIRKRTKHQSFSLAVAGLNTISLGRSLKEISDGVITISGRNLCYANTINW 477
QY 477 DQLEFRPHQALLHTANPEDECVGEGIACHQLCARHCGWGPCTQCNGSQFLRQGEVE 536
DB 478 KKLFGTGNQKTKIMNNAEKDKAVNVHNCPLCSSSGCGWGPBRDCVSGQNVSRGECVE 537
QY 537 ECRVLQGLPREYVNAARCLFCHEGCPQNGSVTCFPEADQCVAAHYKDPPECVARGPS 596
DB 538 KCNLLEBEPREFENSECICQHEPCIPQANNITICTGRGPDNCICQAHYIDGPRCVTKCPA 597
QY 597 GVKPDLVYMPKPEDEGACQCPINCTHSCTVDLDDKCG 636
DB 598 GIMGEENNTL-VMKYADANNVCHLCHANCITYGCAGPGLQCG 636

Search completed: July 22, 2003, 08:59:55
Job time : 53.1263 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:12:49 ; Search time 10.2304 Seconds

(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-59-73-12

Perfect score: 6809

Sequence: 1 MELALCRWGLLLALPPQA.....TRKGTPTAENPEYLGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476326 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6729	98.8	1255	1	ERB2_HUMAN
2	5925	87.0	1257	1	ERB2_RAT
3	5918.5	86.9	1254	1	ERB2_MESAU
4	3115	45.7	1210	1	EGFR_HUMAN
5	3094	45.4	1210	1	EGFR_MOUSE
6	2950.5	43.3	1308	1	ERB4_HUMAN
7	2931	43.0	1308	1	ERB4_RAT
8	2669.5	39.2	1167	1	XMRK_XIPMA
9	2398.5	35.2	1342	1	ERB3_HUMAN
10	2332.5	34.3	1339	1	ERB3_RAT
11	1932	28.4	1426	1	EGFR_DROME
12	1749.5	25.7	634	1	ERBB_ALV
13	1703	25.0	604	1	ERBB_AVIER
14	1630	23.9	540	1	ERBB_AVIER
15	1586	23.3	703	1	EGFR_CHICK
16	1275	18.7	1323	1	LIZ23_CAEBL
17	1142.5	18.8	245	1	ERB2_MOUSE
18	742	10.9	1363	1	ILPR_BRALA
19	715	10.5	1382	1	INSR_HUMAN
20	710	10.4	1383	1	INSR_RAT
21	709.5	10.4	1372	1	INSR_MOUSE
22	703	10.3	1300	1	IRR_MOUSE
23	700	10.3	1607	1	MYR_LYMT
24	694	10.2	1297	1	IRR_HUMAN
25	689.5	10.1	1300	1	IRR_HUMAN
26	672	9.9	1477	1	HTK7_HYDAT
27	653	9.6	1367	1	IGIR_HUMAN
28	643	9.4	1373	1	IGIR_MOUSE
29	639.5	9.4	1370	1	IGIR_RAT
30	635	9.3	1390	1	INSR_AEDAE
31	613	9.0	2146	1	INSR_DROME
32	610	9.0	987	1	EPB4_HUMAN
33	591.5	8.7	984	1	EPB1_CHICK

34	589.5	8.7	977	1	EPB2_MOUSE	003145 mus musculus
35	588	8.6	1114	1	RET_HUMAN	P07949 homo sapien
36	586	8.6	987	1	EPB4_MOUSE	P54761 mus musculus
37	584.5	8.6	976	1	EPB2_HUMAN	P23317 homo sapien
38	583.5	8.6	984	1	EPB1_RAT	P09759 ratius norv
39	579	8.5	902	1	EPBB_XENLA	Q91736 xenopus lae
40	577.5	8.5	984	1	EPB1_HUMAN	P54762 homo sapien
41	576.5	8.5	985	1	EPBA_XENLA	Q91571 xenopus lae
42	573.5	8.4	1053	1	FAK1_CHICK	Q00944 gallus gall
43	569	8.4	1068	1	FAK1_XENLA	Q91738 xenopus lae
44	567	8.3	757	1	HTI6_HYDAT	P53356 hydra atten
45	563	8.3	1052	1	FAK1_MOUSE	P34152 mus musculus

ALIGNMENTS

RESULT 1	ERB2_HUMAN	STANDARD:	PRT: 1255 AA.
ID	ERB2_HUMAN		
AC	P04626;		
DT	13-AUG-1987 (Rel. 05, Created)		
DT	13-AUG-1987 (Rel. 05, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)		
DE	(p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell		
GN	ERBB2 OR HER2 OR NGL OR NEU.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Euteria; Primates; Carnivora; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=86118663; PubMed=3003577;		
RA	Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,		
RA	Saito T., Toyoshima K.,		
RT	"Similarity of protein encoded by the human c-erb-B-2 gene to		
RT	epidermal growth factor receptor.";		
RL	Nature 319:230-234(1986).		
RN	(2)		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=86016729; PubMed=2999374;		
RA	Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,		
RA	McGrath J., Seeburg P.H., Ullrich A., Schlessinger J.,		
RT	"Tyrosine kinase receptor with extensive homology to EGF receptor		
RT	shares chromosomal location with neu oncogene.";		
RL	Science 230:1132-1139(1985).		
RN	(3)		
RP	SEQUENCE OF 737-1031 FROM N.A.		
RX	MEDLINE=86016729; PubMed=2999374;		
RA	Semba K., Kamata N., Toyoshima K., Yamamoto T.,		
RT	"A v-erbB-related protooncogene, c-erbB-2, is distinct from the		
RT	c-erbB-1/epidermal growth factor receptor gene and is amplified in a		
RT	human salivary gland adenocarcinoma.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 82:6437-6501(1985).		
RN	(4)		
RP	VARIANTS VAL-654 AND VAL-655.		
RX	MEDLINE=93194196; PubMed=8095488;		
RA	Ehsani A., Low J., Wallace R.B., Wu A.M.,		
RT	"Characterization of a new allele of the human ERBB2 gene by allele-		
RT	specific competition hybridization.";		
RL	Genomics 15:426-429(1993).		
CC	-!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,		
CC	AUTOPHAG NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A		
CC	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-		
CC	ALPHA AND AMPHIREGULIN.		
CC	-!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein		
CC	tyrosine phosphate.		
CC	-!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS		
CC	(POTENTIAL).		
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.		

CC - P.TM. LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES (BY SIMILARITY).
 CC - POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN
 CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
 CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
 CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M11767; AAA35808.1; JOINED.
 DR EMBL; M11761; AAA35808.1; JOINED.
 DR EMBL; M11762; AAA35808.1; JOINED.
 DR EMBL; M11763; AAA35808.1; JOINED.
 DR EMBL; M11764; AAA35808.1; JOINED.
 DR EMBL; M11765; AAA35808.1; JOINED.
 DR EMBL; M11766; AAA35808.1; JOINED.
 DR EMBL; M11730; AAA75493.1; -
 DR EMBL; M12036; AAA35978.1; -
 DR EMBL; X03363; CAA27060.1; -
 DR PIR; A25491; A25491.
 DR PIR; A24571; A24571.
 DR HSSP; P11362; 1FGK.
 DR Genew; HGNC:3430; ERBB2.
 DR MIM; 164870; -
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_kinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_kinase.
 DR InterPro; IPR004019; YLP_motif.
 DR Pfam; PF00069; kinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR Pfam; PF02757; YLP; 2.
 DR ProDom; PD000001; Euk_kinase; 1.
 DR SMART; SM00261; Fu_3.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KM Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KM Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KM Polymorphism.
 FT SIGNAL 1 21
 FT CHAIN 22 1255
 FT DOMAIN 22 652
 FT TRANSMEM 653 675
 FT DOMAIN 676 1255
 FT NP_BIND 720 987
 FT BINDING 726 734
 FT BINDING 753 753
 FT ACT_SITE 845 845
 FT DISULFID 195 204
 FT DISULFID 199 212
 FT DISULFID 220 227
 FT DISULFID 224 235
 FT DISULFID 236 244
 FT DISULFID 240 252
 FT DISULFID 255 264
 FT DISULFID 268 295
 FT DISULFID 299 311
 FT DISULFID 315 331
 FT DISULFID 334 338
 FT DISULFID 511 520
 FT DISULFID 515 528
 FT DISULFID 531 540
 FT DISULFID 544 560
 FT DISULFID 544 560

FT DISULFID 563 576
 FT DISULFID 567 584
 FT DISULFID 596 596
 FT DISULFID 600 623
 FT DISULFID 626 634
 FT DISULFID 630 642
 FT MOD_RES 1139 1139
 FT MOD_RES 1248 1248
 FT CARBOHYD 68 68
 FT CARBOHYD 124 124
 FT CARBOHYD 187 187
 FT CARBOHYD 259 259
 FT CARBOHYD 530 530
 FT CARBOHYD 571 571
 FT CARBOHYD 629 629
 FT VARIANT 634 654
 FT VARIANT 654 654
 FT VARIANT 655 655
 FT CONFLICT 1170 1170
 FT SEQUENCE 1255 AA; 137909 MW; 39BDPDA04DCF962 CRC64;
 SQ
 Query Match 98.8%; Score 6729; DB 1; Length 1255;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
 QY 1 MELALACRMGLTALLPFGASVTCCTGDMKRLRLPASPTHLMRLHLYOGCCVCGNTL 60
 1 MELALACRMGLTALLPFGASVTCCTGDMKRLRLPASPTHLMRLHLYOGCCVCGNTL 60
 QY 61 ELTYLPTNALSFLDDIOEVQCIYKANSKFGITELQRLRTVGTGLEFEDNYALAVLDNG 120
 61 ELTYLPTNALSFLDDIOEVQCVVLIANQVRQVPLGLRLTVRGTFLEEDNYALAVLDNG 120
 Db 61 ELTYLPTNALSFLDDIOEVQCVVLIANQVRQVPLGLRLTVRGTFLEEDNYALAVLDNG 120
 QY 121 DPLNNTPTPTGASPGSLBELQRLSTELTKGVALIQRPQCYQDTIMKQIFKQNOA 180
 121 DPLNNTPTPTGASPGSLBELQRLSTELTKGVALIQRPQCYQDTIMKQIFKQNOA 180
 Db 121 DPLNNTPTPTGASPGSLBELQRLSTELTKGVALIQRPQCYQDTIMKQIFKQNOA 180
 QY 181 LTLIDNRSRACHPSPCMKSGRSGESSEDCQSLTRVYAGGACGRCGLPTCCHEQC 240
 181 LTLIDNRSRACHPSPCMKSGRSGESSEDCQSLTRVYAGGACGRCGLPTCCHEQC 240
 Db 181 LTLIDNRSRACHPSPCMKSGRSGESSEDCQSLTRVYAGGACGRCGLPTCCHEQC 240
 QY 241 AAGCTGPKSDCLACLPHNSGICELCPALVYNTDTFESMPNEGAYTFGASCVTACP 300
 241 AAGCTGPKSDCLACLPHNSGICELCPALVYNTDTFESMPNEGAYTFGASCVTACP 300
 Db 241 AAGCTGPKSDCLACLPHNSGICELCPALVYNTDTFESMPNEGAYTFGASCVTACP 300
 QY 301 YNYLSDVSGCTLVCPLNQVETAEQDQRECKSKPARVCYGLGMEHLVEVRAVTSAN 360
 301 YNYLSDVSGCTLVCPLNQVETAEQDQRECKSKPARVCYGLGMEHLVEVRAVTSAN 360
 Db 301 YNYLSDVSGCTLVCPLNQVETAEQDQRECKSKPARVCYGLGMEHLVEVRAVTSAN 360
 QY 361 IOEFAGCKKIFGSLAFDPESFDGDSANTAPLOPEQLQVFTLEITGYLYISAMPDLP 420
 361 IOEFAGCKKIFGSLAFDPESFDGDSANTAPLOPEQLQVFTLEITGYLYISAMPDLP 420
 Db 361 IOEFAGCKKIFGSLAFDPESFDGDSANTAPLOPEQLQVFTLEITGYLYISAMPDLP 420
 QY 421 DISVPGNLTQVIRGLHNGAVSLTQIGISMLGRSLRELSGSLALHNHTHLCFPHTV 480
 421 DISVPGNLTQVIRGLHNGAVSLTQIGISMLGRSLRELSGSLALHNHTHLCFPHTV 480
 Db 421 DISVPGNLTQVIRGLHNGAVSLTQIGISMLGRSLRELSGSLALHNHTHLCFPHTV 480
 QY 481 PNDQLFNPHQALLHTANRPEDECEVGEGLAQCLCARGHGCPPTQCVCNSQFLRQEC 540
 481 PNDQLFNPHQALLHTANRPEDECEVGEGLAQCLCARGHGCPPTQCVCNSQFLRQEC 540
 Db 481 PNDQLFNPHQALLHTANRPEDECEVGEGLAQCLCARGHGCPPTQCVCNSQFLRQEC 540
 QY 541 VECRCYLOGLPREYNNANCLPCHECCOPONGSVTCFPEADQVCACAHYDPPFCVARC 600
 541 VECRCYLOGLPREYNNANCLPCHECCOPONGSVTCFPEADQVCACAHYDPPFCVARC 600
 Db 541 VECRCYLOGLPREYNNANCLPCHECCOPONGSVTCFPEADQVCACAHYDPPFCVARC 600
 QY 601 PSQVPRDSYPMIWFPPBEGACQCPINCCHSCVYDDDKCPAEPASPLITSISAVG 660
 601 PSQVPRDSYPMIWFPPBEGACQCPINCCHSCVYDDDKCPAEPASPLITSISAVG 660
 Db 601 PSQVPRDSYPMIWFPPBEGACQCPINCCHSCVYDDDKCPAEPASPLITSISAVG 660
 QY 661 ILLVVLGVVFSIILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNOAMRLKETEL 720
 661 ILLVVLGVVFSIILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNOAMRLKETEL 720
 Db 661 ILLVVLGVVFSIILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNOAMRLKETEL 720


```

FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 626 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

```

```

Query Match 87.0%; Score 5925; DB 1; Length 1257;
Best Local Similarity 87.0%; Pred. No. 9.7e-309;
Matches 1094; Conservative 49; Mismatches 112; Indels 2; Gaps 2;

```

```

QY 1 MELALGRWGLLALPPGAASITQCTGIDMKLRPASPEHLDMLRHHYOGQVQVQNTL 60
DB 1 MELAMCRWGLLALPPGIAAGTQCTGIDMKLRPASPEHLDMLRHHYOGQVQVQNTL 60
QY 61 ELTYLPTNASLSTFQDQEOVQYIKANSKIGITTELORLATVGTOLFEDNTALAVLNG 120
DB 61 ELTYVPANASLSFQDQEOVQYVLAHNOVKRVPQRLATVGTOLFEDNTALAVLNR 120
QY 121 DPLNNTPTVT-GASPGGLRELOLRSLTEILKGGVLIQBNPOLCYODTILMKDIFHKNNOL 179
DB 121 DPONVASTPTGRFPEGLRELOLRSLTEILKGGVLIQBNPOLCYODTILMKDIFHKNNOL 180
QY 180 ALTLIDTNRSHACHPCSPMCKGSRGWGESSEDCQSLTRTVACAGCARCKRPLPTDCHEQ 239
DB 181 APVIDITNRSHACHPCSPACKDNKCHWGESSEDCQILITGITCGACARCKRPLPTDCHEQ 240
QY 240 CAAGCTGPKHSDCLACHFNHSGICEHCALTYNTDFEESNPEEGYTGASCVTAC 299
DB 241 CAAGCTGPKHSDCLACHFNHSGICEHCALTYNTDFEESNPEEGYTGASCVTAC 300
QY 300 PNYLSTDVSGCTLVCPHLNQEVTAEADGTORCEKSKPCARVCYGLGMEHLREVAVTSA 359
DB 301 PNYLSTEVGSCITLVCPNNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVAVTSD 360
QY 360 NIOEFAGCKIFGSLAPESFGDSPASNAPIQPOLQVFEFLTEITGYLYISAMPDSL 419
DB 361 NVQEFDCCKIFGSLAPESFGDSPSGIAPLRPOLQVFEFLTEITGYLYISAMPDSL 420
QY 420 PDLSPQNLQVIRGRIILHNGAYSLTQGLGISWLGRLRELGSGALIHNTHTLCFYHT 479
DB 421 RDLSPQNLRIIRGRIILHNGAYSLTQGLGISWLGRLRELGSGALIHNTHTLCFYHT 480
QY 480 VPMQDLFRNHQALHNAHPEDB-CYBGLACHOLCANGCHGCPPEPTCCVNCOSQPLRQ 538
DB 481 VPMQDLFRNHQALHNGRPEEDLVCSSSLVNSLCAHCHGCPPTCCVNCOSQPLRQ 540
QY 539 ECVEBECRVLOGLPREYVNAHACLPCHEPCOPONGSVTCFEPADQVCAAHKYPFCYA 598
DB 541 ECVEBECRVWKGPREYVSDKCLPCHEPCOPONGSVTCFEPADQVCAAHKYPFCYA 600
QY 599 RCPSGVKRPDLSTYPIKWFPEBEACQPCPINCTHSCVDDDKCQPAORASPLTISAV 658
DB 601 RCPSGVKRPDLSTYPIKWFPEBEACQPCPINCTHSCVDDDKCQPAORASPLTISAV 660
QY 659 VGLLVVVLGVVFGLILKRRQOKIRKTYMRLLQETELVEPLTPSGAMNQOMILKET 718
DB 661 VGLLVVVLGVVFGLILKRRQOKIRKTYMRLLQETELVEPLTPSGAMNQOMILKET 720
QY 719 ELRKVKLGSAGAGTYVKGWIPDGENVKIPVAIKYLRNTSPKANKELLDEAYYMAVG 778
DB 721 ELRKVKLGSAGAGTYVKGWIPDGENVKIPVAIKYLRNTSPKANKELLDEAYYMAVG 780
QY 779-SPVYSRLGICLNSTVGLVQLMPYCGLLDHNRENGRGLSGODLLNWCQIAKMSYLED 838

```

```

DB 781 SPVYSRLGICLNSTVGLVQLMPYCGLLDHNRENGRGLSGODLLNWCQIAKMSYLED 840
QY 839 VRLVHRDLAARNVLVSPNHVKTTFDGLARLLDIDETEHADGKVPYKMALESILRRR 898
DB 841 VRLVHRDLAARNVLVSPNHVKTTFDGLARLLDIDETEHADGKVPYKMALESILRRR 900
QY 899 FTHOSDWSGVVWELMTFGAKPYGIPAREIPDLLEGERLPQPICTIDVYIMVNC 958
DB 901 FTHOSDWSGVVWELMTFGAKPYGIPAREIPDLLEGERLPQPICTIDVYIMVNC 960
QY 959 WMTDSECRPFRELVSFMSMARDPQRFVVIQNEIDGPASPLDSTFYRSLLDEDDMGDLV 101
DB 961 WMTDSECRPFRELVSFMSMARDPQRFVVIQNEIDGPASPLDSTFYRSLLDEDDMGDLV 102
QY 1019 DAEHYLYPQGFPCPDPAAGAMVHRRSSSTRSGGGLTLGLEPSEEAERSLAAS 107
DB 1021 DAEHYLYPQGFPCPDPAAGAMVHRRSSSTRSGGGLTLGLEPSEEAERSLAAS 108
QY 1079 EGAGSDVFDGDLGMAAKGLQSLPTHDPEPLQRYSEDPTVPLPSSTDGYVAPLTCSPQPE 113
DB 1081 EGAGSDVFDGDLGMAAKGLQSLPTHDPEPLQRYSEDPTVPLPSSTDGYVAPLTCSPQPE 114
QY 1139 YVNGPDYRPPPSREGPLPAPAPAGATLERAXTSPGXGVVQVFAFGAVENPEYLT 119
DB 1141 YVNGSEVQPPPLTPREGPLPAPAPAGATLERAXTSPGXGVVQVFAFGAVENPEYLT 120
QY 1199 PGGGAAPQHPAPAPSPAPNLYVMODPREGAPSTKGTPTMENPEYLGADV 1255
DB 1201 PGGGAAPQHPAPAPSPAPNLYVMODPREGAPSTKGTPTMENPEYLGADV 1257

```

```

RESULT 3
ERR2_MESAU STANDARD; PRT; 1254 AA.
AC 06053;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase etbb-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
OS ERBB2 OR NEU
NC Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94193007; PubMed=7908275;
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255 (1994).
CC -1- FUNCTION. ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```


CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL, D16295; BAA03801.1; --

DR HSP, P11362; IFGK.

DR InterPro: IPR000494; EGFR_L_domain.

DR InterPro: IPR000719; Euk_kinase.

DR InterPro: IPR002174; Furin-like.

DR InterPro: IPR001245; Tyr_kinase.

DR InterPro: IPR004019; YLP_motif.

DR Pfam: PF00065; Pkinase.1.

DR Pfam: PF00757; Furin-like.1.

DR Pfam: PF01030; Recep_L_domain.2.

DR Pfam: PF02757; YLP.2.

DR ProDom: PD000001; Euk_kinase.1.

DR SMART: SM00261; Fu.3.

DR SMART: SM00219; Tyrc.1.

DR PROSITE: PS00107; PROTEIN_KINASE_ATP.1.

DR PROSITE: PS00109; PROTEIN_KINASE_TIR.1.

DR PROSITE: PS50011; PROTEIN_KINASE_DOM.1.

DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;

KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;

KW Proto-oncogene; Disease mutation.

FT SIGNAL 1 21

FT CHAIN 22 1254

FT DOMAIN 22 652

FT TRANSSEM 653 675

FT DOMAIN 676 1254

FT DOMAIN 158 368

FT DOMAIN 472 644

FT DOMAIN 720 987

FT NP_BIND 726 734

FT BINDING 753 753

FT ACT_SITE 845 845

FT DISULFID 195 204

FT DISULFID 199 212

FT DISULFID 236 244

FT DISULFID 240 252

FT DISULFID 285 264

FT DISULFID 268 295

FT DISULFID 299 311

FT DISULFID 315 331

FT DISULFID 334 338

FT DISULFID 511 520

FT DISULFID 515 528

FT DISULFID 531 540

FT DISULFID 544 560

FT DISULFID 563 576

FT DISULFID 567 584

FT DISULFID 587 596

FT DISULFID 600 623

FT DISULFID 626 634

FT DISULFID 630 642

FT MOD_RES 1139 1139

FT MOD_RES 1247 1247

FT CARBOHYD 68 68

FT CARBOHYD 125 125

FT CARBOHYD 187 187

FT CARBOHYD 259 259

FT CARBOHYD 530 530

FT CARBOHYD 571 571

FT CARBOHYD 629 629

FT CARBOHYD 658 658

FT VARIANT 659 659

SO SEQUENCE 1254 AA; 138252 MW; 974C3791C212B2B1 CRC64;

Query Match 86.98; Score 5918.5; DB 1; Length 1254;

Best Local Similarity 86.88; Pred. No. 2.2e-108;

Matches 1089; Conservative 58; Mismatches 107; Indels 1; Gaps 1;

Db 1 MELALCRWGILLALLPPGAASTOVCTGTDMLRLPASBETHLDMRLHYOCGVVQGNL 60

Qy 61 ELTYLPNTASLSFLQDIOEVOYIKANSKFIGITEHORLRIYRGTOLEFDNYALALNDG 120

Db 61 ELTYLPNTASLSFLQDIOEVOYIKANSKFIGITEHORLRIYRGTOLEFDNYALALNDG 120

Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQNFOLCYODTILMKDI FHKNOQLA 180

Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQNFOLCYODTILMKDI FHKNOQLA 180

Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQNFOLCYODTILMKDI FHKNOQLA 180

Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQNFOLCYODTILMKDI FHKNOQLA 180

Qy 181 LTLIDNRSRACHPCSPMKSGKSCWSESDOSLTRYCAGGACRCKPLPTDCCHEOC 240

Db 181 LTLIDNRSRACHPCSPMKSGKSCWSESDOSLTRYCAGGACRCKPLPTDCCHEOC 240

Qy 181 LTLIDNRSRACHPCSPMKSGKSCWSESDOSLTRYCAGGACRCKPLPTDCCHEOC 240

Db 181 LTLIDNRSRACHPCSPMKSGKSCWSESDOSLTRYCAGGACRCKPLPTDCCHEOC 240

Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALTYNTDTESMPNEGRTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALTYNTDTESMPNEGRTFGASCVTACP 300

Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALTYNTDTESMPNEGRTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALTYNTDTESMPNEGRTFGASCVTACP 300

Qy 301 YNYLSDVGSCTLVCTLANOEVTAEAGTORCEKSKPCARVICYGLAMEHLREYAVTSAN 360

Db 301 YNYLSDVGSCTLVCTLANOEVTAEAGTORCEKSKPCARVICYGLAMEHLREYAVTSAN 360

Qy 301 YNYLSDVGSCTLVCTLANOEVTAEAGTORCEKSKPCARVICYGLAMEHLREYAVTSAN 360

Db 301 YNYLSDVGSCTLVCTLANOEVTAEAGTORCEKSKPCARVICYGLAMEHLREYAVTSAN 360

Qy 361 IOEPAGCKKIFGSLAPLPSFDGDPASNTAPLOPELOVFEETLEETGYLY:SANPDSL 420

Db 361 IOEPAGCKKIFGSLAPLPSFDGDPASNTAPLOPELOVFEETLEETGYLY:SANPDSL 420

Qy 361 IOEPAGCKKIFGSLAPLPSFDGDPASNTAPLOPELOVFEETLEETGYLY:SANPDSL 420

Db 361 IOEPAGCKKIFGSLAPLPSFDGDPASNTAPLOPELOVFEETLEETGYLY:SANPDSL 420

Qy 421 DLSVPQNTQYIRGRILHNGAYSLTLQGLGISWLGRLSRLSGLALIHNTHLGCVHTV 480

Db 421 DLSVPQNTQYIRGRILHNGAYSLTLQGLGISWLGRLSRLSGLALIHNTHLGCVHTV 480

Qy 421 DLSVPQNTQYIRGRILHNGAYSLTLQGLGISWLGRLSRLSGLALIHNTHLGCVHTV 480

Db 421 DLSVPQNTQYIRGRILHNGAYSLTLQGLGISWLGRLSRLSGLALIHNTHLGCVHTV 480

Qy 481 PMQOLFNPQALLHNPANRDEDCVEGLACHQLCARGHCMPGPGQCNCGQFLRGCC 540

Db 481 PMQOLFNPQALLHNPANRDEDCVEGLACHQLCARGHCMPGPGQCNCGQFLRGCC 540

Qy 481 PMQOLFNPQALLHNPANRDEDCVEGLACHQLCARGHCMPGPGQCNCGQFLRGCC 540

Db 481 PMQOLFNPQALLHNPANRDEDCVEGLACHQLCARGHCMPGPGQCNCGQFLRGCC 540

Qy 541 VEECRVLOGLPREYVABHCLPCHPEPCQONSQVTCFGEADQCVACAHYKPPFCVAC 600

Db 541 VEECRVLOGLPREYVABHCLPCHPEPCQONSQVTCFGEADQCVACAHYKPPFCVAC 600

Qy 541 VEECRVLOGLPREYVABHCLPCHPEPCQONSQVTCFGEADQCVACAHYKPPFCVAC 600

Db 541 VEECRVLOGLPREYVABHCLPCHPEPCQONSQVTCFGEADQCVACAHYKPPFCVAC 600

Qy 541 VKCRVWKGKLPREYVABHCLPCHPEPCQONSQVTCFGEADQCVACAHYKPPFCVAC 600

Db 541 VKCRVWKGKLPREYVABHCLPCHPEPCQONSQVTCFGEADQCVACAHYKPPFCVAC 600

Qy 601 PSQVPLSLMPKPPDEBAGCQPCINCTHSQVLDLCKGCPAEGRASPLTSIVAVG 660

Db 601 PSQVPLSLMPKPPDEBAGCQPCINCTHSQVLDLCKGCPAEGRASPLTSIVAVG 660

Qy 601 PSQVPLSLMPKPPDEBAGCQPCINCTHSQVLDLCKGCPAEGRASPLTSIVAVG 660

Db 601 PSQVPLSLMPKPPDEBAGCQPCINCTHSQVLDLCKGCPAEGRASPLTSIVAVG 660

Qy 661 ILLVVLGVVFGILIRPOQKIRKYMRLLOETLEVEPLTSGAMPNOAKRILKETEL 720

Db 661 ILLVVLGVVFGILIRPOQKIRKYMRLLOETLEVEPLTSGAMPNOAKRILKETEL 720

Qy 661 ILLVVLGVVFGILIRPOQKIRKYMRLLOETLEVEPLTSGAMPNOAKRILKETEL 720

Db 661 ILLVVLGVVFGILIRPOQKIRKYMRLLOETLEVEPLTSGAMPNOAKRILKETEL 720

Qy 721 RYKVLGSGAFGVYKGIWIPDGENVKIPVALKVLRENTSPANKELIDEAYVAVGSP 780

Db 721 RYKVLGSGAFGVYKGIWIPDGENVKIPVALKVLRENTSPANKELIDEAYVAVGSP 780

Qy 721 RYKVLGSGAFGVYKGIWIPDGENVKIPVALKVLRENTSPANKELIDEAYVAVGSP 780

Db 721 RYKVLGSGAFGVYKGIWIPDGENVKIPVALKVLRENTSPANKELIDEAYVAVGSP 780

Qy 781 YVERLLGICLTSTVOQLTQMPGCLLDHVRHGRGLSGQDLNMCVQIAKMSYLEV 840

Db 781 YVERLLGICLTSTVOQLTQMPGCLLDHVRHGRGLSGQDLNMCVQIAKMSYLEV 840

Qy 781 YVERLLGICLTSTVOQLTQMPGCLLDHVRHGRGLSGQDLNMCVQIAKMSYLEV 840

Db 781 YVERLLGICLTSTVOQLTQMPGCLLDHVRHGRGLSGQDLNMCVQIAKMSYLEV 840

Qy 841 LVHRDLAARNVLVKSNHVKITDFGLARLLDDETEYHADGKVPKIMVALSIRRRFT 900

Db 841 LVHRDLAARNVLVKSNHVKITDFGLARLLDDETEYHADGKVPKIMVALSIRRRFT 900

Qy 841 LVHRDLAARNVLVKSNHVKITDFGLARLLDDETEYHADGKVPKIMVALSIRRRFT 900

Db 841 LVHRDLAARNVLVKSNHVKITDFGLARLLDDETEYHADGKVPKIMVALSIRRRFT 900

Qy 901 HOSDWSYGVTTWELMTFGAKPYDGIIPAEIPDLLEKGRLLPQPICTIDVYIMVKCM 960

Db 901 HOSDWSYGVTTWELMTFGAKPYDGIIPAEIPDLLEKGRLLPQPICTIDVYIMVKCM 960

Qy 901 HOSDWSYGVTTWELMTFGAKPYDGIIPAEIPDLLEKGRLLPQPICTIDVYIMVKCM 960

Db 901 HOSDWSYGVTTWELMTFGAKPYDGIIPAEIPDLLEKGRLLPQPICTIDVYIMVKCM 960

Qy 961 IDSECRPRPELSEESRMAKRDQRFVIVQNEDELGPASPLDSTFYRSLLEDMDMDLVDA 102

Db 961 IDSECRPRPELSEESRMAKRDQRFVIVQNEDELGPASPLDSTFYRSLLEDMDMDLVDA 102

Qy 961 IDSECRPRPELSEESRMAKRDQRFVIVQNEDELGPASPLDSTFYRSLLEDMDMDLVDA 102

Db 961 IDSECRPRPELSEESRMAKRDQRFVIVQNEDELGPASPLDSTFYRSLLEDMDMDLVDA 102

Qy 1021 EBYLVPQGGFFCPDPAAGAGMYHHRHRSSTRSGGDLTLGLBSEEBEAPRSPAPSEB 108

Db 1021 EBYLVPQGGFFCPDPAAGAGMYHHRHRSSTRSGGDLTLGLBSEEBEAPRSPAPSEB 108

Qy 1021 EBYLVPQGGFFCPDPAAGAGMYHHRHRSSTRSGGDLTLGLBSEEBEAPRSPAPSEB 108

Db 1021 EBYLVPQGGFFCPDPAAGAGMYHHRHRSSTRSGGDLTLGLBSEEBEAPRSPAPSEB 108

Qy 1081 AGSDVDFDGLGMAAGLQSLPHHDSPLQKRSDEPTVPLPSETGTGYVAPLTCSPQPEV 114

Db 1081 AGSDVDFDGLGMAAGLQSLPHHDSPLQKRSDEPTVPLPSETGTGYVAPLTCSPQPEV 114

```

Cy 1141 NOPDVPRQPPSPREGULPAPARPGATLEPAKTLSPGKNGVWQVFAFGAVENPEYLTPO 1200
    |||:|||||
Db 1141 NQPEVRQPPLEPEGLPEVPRPAGATLEPKTLSPGKNGVWQVDFTFGGAENPEYLVPR 1200
    |||:|||||
Cy 1201 GGAAPCPHPPPPSPAFNDLYWDDQDPERGAPEPSPFKGTPAENPEYGLDVPV 1255
    |||:|||||
Db 1201 GGSASQPH-PPALCPAFNDLYWDDQDPERGSPPNPFESTPTAENPEYGLDVPV 1254
    |||:|||||

RESULT 4
EGFR HUMAN
ID EGFR_HUMAN STANDARD; PRT; 1210 AA.
AC P00533; P06268; Q14225; Q9UMD8; Q9UMG5; Q92795; Q00732;
AC Q00688; Q9B282; Q9H2C9; Q9GZK1; Q9H3C9;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
DE protein-tyrosine kinase ErbB-1).
GN EGFR; OR ERBB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=84219729; PubMed=6328312;
RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
RA Maves E.L.V., Whitte N., Waterfield M.D., Seeburg P.R.;
RT "Human epidermal growth factor receptor cDNA sequence and aberrant
RT expression of the amplified gene in A431 epidermoid carcinoma cells.";
RL Nature 309:418-425(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX TISSUE=Placenta;
RX MEDLINE=95382937; PubMed=7654368;
RA Ilekis J.V., Stark B.C., Scoccia B.;
RT "Possible role of variant RNA transcripts in the regulation of
RT epidermal growth factor receptor expression in human placenta.";
RL Mol. Reprod. Dev. 41:149-156(1995).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX TISSUE=Placenta;
RX MEDLINE=9707686; PubMed=8918811;
RA Reiter J.L., Maible N.J.;
RT "A 1.8 kb alternative transcript from the human epidermal growth
RT factor receptor gene encodes a truncated form of the receptor.";
RL Nucleic Acids Res. 24:4050-4056(1996).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX TISSUE=Placenta;
RX MEDLINE=97256547; PubMed=9103388;
RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
RT "Expression of a truncated epidermal growth factor receptor-like
RT protein (TEGFR) in ovarian cancer.";
RL Gynecol. Oncol. 65:36-41(1997).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
RX TISSUE=Placenta;
RX MEDLINE=97256547; PubMed=9103388;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Scheel Sinclair C., Pearse R.S., Green P.J., Yee D., Lampfand A.L.,
RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative EGFR transcripts encoding truncated receptor
RT isoforms.";
RL Genomics 71:1-20(2001).
RN [6]
RP SEQUENCE OF 575-687 FROM N.A.
RA Reiter J.L., Threadgill D.W., Danielson A.J., Scheel C.M.,
RA Lampfand A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maible N.J.;
RT "Human and mouse alternative EGFR transcripts encoding only the
RT extracellular domain of the receptor.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 713-924 FROM N.A.
RX MEDLINE=84196372; PubMed=6332621;
RA Lin C.R., Chen W.S., Krueger W., Stolarzky L.S., Weber W.,
RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
RT "Expression cloning of human EGF receptor complementary DNA: gene
RT amplification and three related messenger RNA products in A431
RT cells.";
RL Science 224:843-848(1984).
RN [8]
RP SEQUENCE OF 150-962 FROM N.A.
RX MEDLINE=84245835; PubMed=6330563;
RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
RA Roe B.A., Merlino G.T., Pastan I.;
RT "Human epidermal growth factor receptor cDNA is homologous to a
RT variety of RNAs overproduced in A431 carcinoma cells.";
RL Nature 309:806-810(1984).
RN [9]
RP SEQUENCE OF 1028-1210 FROM N.A.
RX MEDLINE=85046483; PubMed=6093780;
RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
RA O'Malley B.W.;
RT "Isolation of an evolutionarily conserved epidermal growth factor
RT receptor cDNA from human A431 carcinoma cells.";
RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
RN [10]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=88217333; PubMed=3329716;
RA Haley J.D., Whitte N., Bennett P., Kinchington D., Ullrich A.,
RA Waterfield M.D.;
RT "The human EGF receptor gene: structure of the 110 kb locus and
RT identification of sequences regulating its transcription.";
RL Oncogene Res. 1:375-396(1987).
RN [11]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=91107677; PubMed=1988448;
RA Haley J.D., Waterfield M.D.;
RT "Contributory effects of de novo transcription and premature
RT transcript termination in the regulation of human epidermal growth
RT factor receptor proto-oncogene RNA synthesis.";
RL J. Biol. Chem. 266:1746-1753(1991).
RN [12]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=85270438; PubMed=2991899;
RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
RT "Characterization and sequence of the promoter region of the human
RT epidermal growth factor receptor gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
RN [13]
RP SEQUENCE OF 540.
RX Kohda D.;
RT Submitted (SEP-1997) to the SWISS-PROT data bank.
RN [14]
RP RECEPTOR ACTIVITY.
RX MEDLINE=84191554; PubMed=6325948;
RX Mroczkowski B., Mosig G., Cohen S.;
RT "ATP-stimulated interaction between epidermal growth factor receptor
RT and supercoiled DNA.";
RL Nature 309:270-273(1984).
RN [15]
RP PHOSPHORYLATION.
RX MEDLINE=8927817; PubMed=2543678;
RX Margolis B.L., Lax I., Kriz R., Dombalagian M., Honegger A.M.,
RX Hawk R., Givol D., Ullrich A., Schlessinger J.;
RT "All autophosphorylation sites of epidermal growth factor (EGF)
RT receptor and HER2/neu are located in their carboxyl-terminal tails.
RT Identification of a novel site in EGF receptor.";
RL J. Biol. Chem. 264:10667-10671(1989).
RN [16]

```

RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
 RP ASN-528.
 RX MEDLINE=96398132; PubMed=8962717;
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
 RT "Analysis of the glycosylation patterns of the extracellular domain of
 RT the epidermal growth factor receptor expressed in Chinese hamster
 RT ovary fibroblasts.";
 RL Growth Factors 13:121-132(1996).
 RN (17)
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
 RP ASN-603.
 RX MEDLINE=20198209; PubMed=10731668;
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
 RT "Characterization of the N-oligosaccharides attached to the atypical
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor
 RT receptor.";
 RL J. Biochem. 127:65-72(2000).
 RN (18)
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE=98225196; PubMed=9556602;
 RA Abe Y., Odaka M., Inagaki F., Lax I., Schlessinger J., Kohda D.;
 RT "Disulfide bond structure of human epidermal growth factor receptor.";
 RL J. Biol. Chem. 273:11150-11157(1998).
 RN (19)
 RP REVIEW.
 RX MEDLINE=87297456; PubMed=3039909;
 RA Carpenter G.;
 RT "Receptors for epidermal growth factor and other polypeptide
 RT mitogens.";
 RL Annu. Rev. Biochem. 56:881-914(1987).
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
 CC EGF-like growth factor, GP30 and vaccinia virus growth factor. Is
 CC involved in the control of cell growth and differentiation.
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
 CC secreted.
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms, 1/p170 (shown here), 2/p60/
 CC truncated isoform/TEGFR, 3/p110 and 4; are produced by
 CC alternative splicing.
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
 CC expressed in ovarian cancers.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X00588; CAA25240.1; -
 DR EMBL: U95089; AAB53063.1; -
 DR EMBL: U48722; AAC50802.1; -
 DR EMBL: U48723; AAC50804.1; -
 DR EMBL: U48724; AAC50796.1; -
 DR EMBL: U48725; AAC50797.1; -
 DR EMBL: U48726; AAC50798.1; -

Query Match 45.7% Score 3115; DB 1; Length 1210;
 Best Local Similarity 49.2% Pred. No. 6.9e-159;
 Matches 622; Conservative 177; Mismatches 360; Indels 106; Gaps 21;

QY 11 LLLALLPPGAA--TQVCTGTDMKRLPASPETHLDMLRHLVYGCQGVVQGNLLTYLPTN 68
 DB 14 LLAALCPASRALLEKKYCGGTSTNLTQLGTGFEDHFLSLQKMFNKNCEVVLGNLEITTVQRN 73

QY 69 ASLSPLODIOEVQVQIKANSKRTIGTELORLRIVRGTOLEEDNVALAVLNDGDLPLNTTP 128
 DB 74 YDLSFKTIQEVAGVYLAALNTVEITPLENIQILKGNYYENSTALAVLSYD----- 126
 QY 129 VTGASPGRLREQLRLSTLEIKGVLYLQRPOLCYODTILWKDIFKKNQALTLIDTNR 188
 DB 127 ---ANKTGLKEIPKMNLOELHGAARFENNPLCNVESIQWRDIYSDFLELMSMDPQNH 183
 QY 189 SAACHPCSMCKGSGCKWSESSDQSLRTYCAGCA-RCKGRLPTDCCHQCAAGCTGP 247
 DB 184 LGSQCKQCPSCPNKSGCKWAGEENCKKIKIICAQCCSRCKGKSPSCCHQCAAGCTGP 243
 QY 248 KASDCLACLFHNSHICELCPALVTYNTDFESPNBEGYTPGASCVTCAPNYLSTD 307
 DB 244 RESDCLVCKKFDKATCDTCEPLMLTYPTTYQMDVNEGRKTSGATVKKCPKNYVTD 303
 QY 308 VSSCTLVCPHNOEVTADDTQRCCKSKPCARVCYIGMEHLREVAVTSANIQEPAGC 367
 DB 304 HSGCVACGASBYEM-EDGVKCKCKGCPKRYCNIGIGEFKDSLINATNIKHFYKC 362
 QY 368 KKIIFGSLAFLESPFGDPAASNTARLPQEQLOVFETLEITGYLISAMPDLPPLSVQN 427
 DB 363 TSIQSDHLIPVAFRGDSFTHTPPLDPELDLITKVEITGFLIQAMPENRTDLHAFEN 422
 QY 428 LQVIRGRLHNGAVSLTQGLGISWLGRLSRLRELSGLALIHNTLCEVHTVPMDQLFR 487
 DB 422 LRIIRGTRKQGFSLAVSLNTISLGRSLKESLDGVIIISGNKNCVANTIMWKTLFG 482
 QY 488 NHHQALLTARPEDEYCEGLACHQLCARHCHGPGPTQCVNCSQFLRGCEVEEVRVL 547
 DB 483 TSGQTKTISNNGEASCATQVCHALCSPEGCSPERPDVSCRNVSRGRECVDKNLL 542
 QY 548 QGLPREYVNAHCLPCHPEQCPONGSVTCFPEADQVACAHYDPPFCVARGCSGVKPD 607
 DB 543 EEPREPEFNSECTIOCHECLPQAMNITTCRGDNICQAHYIDGHCYKTCAGWVGE 602
 QY 608 LSYMPIWKPPEEGACQCPINCHSCVDLDKCKPAEORASPLTSIVSAVG---ILLV 664
 DB 603 NNTL-VKRYADAGHCHLCHNCTGCTGCPLESCPTNGKPIF--SIATGVGALLLV 659
 QY 665 VLVGVFGLILKRPQKIRKRYTMRLELOETVELPLTPSGAMPQAOIRLKEITELKVK 724
 DB 660 VALGIG--LNRHRHIVKRLRLLOERVELPLTPSGAPQALLRLKEIEFFKIK 716
 QY 725 VLGSGAFGVYKGIWIDGENVKIPVAKVLRNTSPKANKEILDEAYVWAGVSPVSR 784
 DB 717 VLGSGAFGVYKGIWIDGENVKIPVAKVLRNTSPKANKEILDEAYVWASVDNPRVCR 776
 QY 785 LLGILSTVOLVQMLPFGCLLDHVENRGLASQDLNWCQIAKMSYLEDVRLVHR 844
 DB 777 LLGILSTVOLVQMLPFGCLLDHVENRGLASQDLNWCQIAKMSYLEDVRLVHR 836
 QY 845 DLAAARNVVKPKPNVKTIDFGLARLLIDETEVYADGKVPIMMMALESLRRFRTHQSD 904
 DB 837 DLAAARNVVKTPQVHKITIDFGLARLLIDETEVYADGKVPIMMMALESLRRFRTHQSD 896
 QY 905 VMSYGVTVWELMTGARPYDGIIPAREIPDLLEKGBRLPOPCTIDVYIMVKNMIDSE 964
 DB 897 VMSYGVTVWELMTGARPYDGIIPAREIPDLLEKGBRLPOPCTIDVYIMVKNMIDSE 956
 QY 965 CRPRPRELVSEFSMARSDPQRFVITQ-NEDLGASPLDSIFYSLSLEDDMGVLVAEY 1023
 DB 957 SRPKRELILFESMAADPQRYVLTQGDERRHLPSPTDSFYALWDEEDMDVVVADEV 1016
 QY 1024 LVPOQGFCDPDPAPGAGVHHRRSSSTRSGGADLLTGLPESEEAAPRSLAPSGAGS 1083
 DB 1017 LVPOQGF-----SSPSKRPFLSLSSAIS 1042
 QY 1084 DVFGDDLCMAKAKLQGLPHTDPSLQRYSEDPTVPFST--DGVAAPLTCSPOPEYVN 1141
 DB 1043 N--NSTVACIDRNLQSCPTKEDSFQORYSSDPTGALTEDSIDDTPL-----PVPEYIN 1094

```

QY 1142 QPDRPQPPSPRECEUPAARAPAGATLERAKTISPGKGVYKDVAFAGAVENPEYL-TPQ 1200
DB 1095 Q-SVPRKPPAGSVONPYVHNPPLNP-----APSRDPRYQD--PSTAVGNPELNTVQ 1143
QY 1201 GGAAPQHPPEPAPFAPDNLVYWDQ-----DP-----PERGAPPSFTFKCTPTAE 1244
DB 1144 -----PTCVNSTFTDSPAHAHQSGHSQISIDNDVDYQQDFPEAKPNKGFKGS-TAE 1193
QY 1245 NPEYL 1249
DB 1194 NAEYL 1198

RESULT 5
EGFR_MOUSE
ID EGFR_MOUSE STANDARD; PRT; 1210 AA.
AC Q01279;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=93026370; PubMed=1408137;
RA Avivi A., Skorecki K., Yayon A., Givol D.;
RT "Promoter region of the murine fibroblast growth factor receptor 2
RT (bek/KGFR) gene.";
RL Oncogene 7:1957-1962(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c, and CD-1; TISSUE=Liver, and Decidua;
RX MEDLINE=93126380; PubMed=7678346;
RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT "Expression of the epidermal growth factor receptor gene is regulated
RT in mouse blastocysts during delayed implantation.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RA Hibbs W.L.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/C3; TISSUE=Liver;
RX MEDLINE=94170986; PubMed=8125255;
RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Barp H.S.,
RA Jenkins N.A., Lee D.C.;
RT "The mouse waved-2 phenotype results from a point mutation in the EGF
RT receptor tyrosine kinase.";
RL Genes Dev. 8:399-413(1994).
RN [5]
RP SEQUENCE OF 1-714 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91232866; PubMed=2030916;
RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;
RT "Comparison of EGF receptor sequences as a guide to study the ligand
RT binding site.";
RL Oncogene 6:673-676(1991).
RN [6]
RP SEQUENCE OF 969-1117 FROM N.A.
RC STRAIN=C3H;
RA Eslinger D.P., Serrero G.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: X78987; CAA5587.1; -
CC EMBL: U03425; AAA17899.1; -
CC EMBL: X59698; CAA42219.1; -
CC EMBL: L06864; AAA53029.1; -
CC EMBL: Z12608; CAA78249.1; -
CC HSSP: P11362; 1FGK.
CC MCD; MG195284; Egfr.
CC InterPro: IPR000494; EGFR_L_domain.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR002174; Furin-like.
CC InterPro: IPR001245; Tyr_pkinase.
CC Pfam: PF00069; Pkinase; 1.
CC Pfam: PF00757; Furin-like; 1.
CC Pfam: PF01030; Recep_L_domain; 2.
CC ProDom: PD000001; Euk_pkinase; 1.
CC SMART; SM00219; Tyrc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE; PS00109; PROTEIN KINASE TYR; 1.
CC Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
CC Tyrosine-protein kinase; Atp-binding; Phosphorylation; Repeat.
CC SGNL 1 24
CC CHAIN 25 1210
CC DOMAIN 25 647
CC TRANSMEM 648 670
CC DOMAIN 671 1210
CC REPEAT 75 300
CC REPEAT 390 600
CC DOMAIN 1028 1071
CC NP_BIND 714 981
CC BINDING 720 728
CC ACT_SITE 839 839
CC DISULFID 190 199
CC DISULFID 194 207
CC DISULFID 215 223
CC DISULFID 219 231
CC DISULFID 232 240
CC DISULFID 236 248
CC DISULFID 251 260
CC DISULFID 264 291
CC DISULFID 295 307
CC DISULFID 311 326
CC DISULFID 329 333
CC DISULFID 506 515
CC DISULFID 510 523
CC DISULFID 526 535
CC DISULFID 539 555
CC DISULFID 558 571
CC DISULFID 582 591
CC DISULFID 595 617
CC DISULFID 620 636
CC MOD_RES 680 680
CC MOD_RES 1092 1092
CC
CC PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
CC PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

```

```

FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY)
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY)
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE)
FT CARBOHYD 128 128 N-LINKED (GLCNAC...) (POTENTIAL)
FT CARBOHYD 175 175 N-LINKED (GLCNAC...) (POTENTIAL)
FT CARBOHYD 196 196 N-LINKED (GLCNAC...) (POTENTIAL)
FT CARBOHYD 352 352 N-LINKED (GLCNAC...) (POTENTIAL)
FT CARBOHYD 413 413 N-LINKED (GLCNAC...) (POTENTIAL)
FT CARBOHYD 444 444 N-LINKED (GLCNAC...) (POTENTIAL)
FT CARBOHYD 528 528 N-LINKED (GLCNAC...) (POTENTIAL)
FT CARBOHYD 568 568 N-LINKED (GLCNAC...) (POTENTIAL)
FT CARBOHYD 603 603 N-LINKED (GLCNAC...) (POTENTIAL)
FT CARBOHYD 623 623 N-LINKED (GLCNAC...) (POTENTIAL)
FT CONFLICT 19 19 C -> S (IN REF. 2)
FT CONFLICT 539 539 C -> W (IN REF. 5)
FT CONFLICT 991 991 L -> F (IN REF. 4)
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6)
SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46DFD2F5 CRC64;

```

Query Match Best Local Similarity 45.4%; Score 3094; DB 1; Length 1210; Pred. No. 9,1e-158; Matches 625; Conservative 170; Mismatches 367; Indels 110; Gaps 23;

```

QY 11 LLLALLPPGAA--STGYCTGTDMKRLRPASPERTHMLRHLVGGCOVQGNLELYLPTN 68
DB 14 LITLCAAGALEKXKVCQSTSNRLTQGLTFEHHFSLQAMYNCEVAGNLEITYQRN 73
QY 69 ASLSFLQDIOEVOOYIKANSKFGITELQRLIRVGTQLEFEDNYALAVLNDGPLNNTTP 128
DB 74 YDLSEFLKTIQEVAGYVIALNLTVERIPLENLQIIRGNALYENTYALATLSN----- 124
QY 129 VTGASPCGLRELRSLTETLKGVLIQNPOLCYODTLMDI----PHKNQALTLI 184
DB 125 -YGINRTGLRELPMKNIQELICAVRFNSNPILCMNDTIQMRDVIQNVFMSMSMDL--- 180
QY 185 DTNRSPACHPCSPWCKSGRCWGBSSDCSLTRTVAGGCA-RCKGPLPTDCCHQCAAG 243
DB 181 -QSHSPSCPKCPCSPGSCWGCSENCCKLTITICAQOCSHRGGRSSDCCHQCAAG 239
QY 244 CTGPKXSDCLACHFNHSGICELHCPALTYNTDTFESPNNEGRTTGASCVTACPNY 303
DB 240 CTGPRESDCLVCKCFDEATCKDTCPLMLYNPTTYQMDVNBEGYXSFATCCKKCPNRY 299
QY 304 LSTDVSGCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYCGMEHLEVRATVSAIOE 363
DB 300 VVTDHGSVVRACGPDIYEV-EDDGIRKCKKCGPCRKVNGIGIEFKDTLSINATNICK 358
QY 364 FAGCKKIFGSLAFLEPSFGDPASNTAPLQEPOLQVFEETIEITGYLYISAMPDLPDS 423
DB 359 FKYCTAISGDLHLPLPAFKGDSFTTRTPPLDPRELEILTKVKEITGFLLIQAMPDMDLH 418
QY 424 VFQNLQVIRGRIIHNQVSLTQGLISLGRSLRELISGSLALHNTHLCFHTVWMD 483
DB 419 AFENLEIIIRGTQHOFSLAVAGLNTISLGRSLKEISDGVIIISGRNLCYANTIMWK 478
QY 484 QLEFNPHQALLHTANPEDEVCEGGLACHQLCARHCWGPFGFTQCVNCSQFLRQGEVEE 543
DB 479 KLFGTPNCKTKIMNNAEKDCAVNVHVCNPLCSSBGWGPBERDVCSONVSRGECVEK 538
QY 544 CRVLOGLPREYVNAHRCLEPCEQCPQNSVTCFEPREADQCYACHYKDPFFCYARCPSG 603
DB 539 CNLIESEPREFENSCICQHPCLPQAMNITCTRGDNCTCQCHYIDGPHCVKCTCPAG 598
QY 604 VKPDISYMPDKFPDEGACQPCPINCTHSCVDLDDKCPAEQRASPLTISVSAVGLL 663
DB 599 IMGENNTL-VMKYADANNVCHLCNANCYTGACGPELQGCCEVWSPSPKPSIATIGVGLL 657
QY 664 VVVLGVVFGI-LIKRQCKIRKTYTRRLLOEVELVEPLTSGAMPNQOMELKTELK 722
DB 658 FIVV-VALSIGLPMKRRIIVKRTLRLLQERELVEPLTSGEAPNQAHLRLKETEFLK 716
QY 723 VVVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAVGSPY 762

```

```

DC 717 IKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAVGSPY 776
QY 783 SRLIGICTSTVQVLVTOLMPYGLLDLHRENRRLSGODLLNMCQIAGMSYLEVRV 842
DB 777 CRLLIGICTSTVQVLVTOLMPYGLLDLHRENRRLSGODLLNMCQIAGMSYLEVRV 836
QY 843 HRDLAARNVYKSNHXYITDFGLARLDDIETRYHADGKVPIMKMALESLIRRFTH 902
DB 837 HRDLAARNVYKSNHXYITDFGLARLDDIETRYHADGKVPIMKMALESLIRRYTH 896
QY 903 SDVMSYVTVWELMTFGAKPYDGIAPREIPDLLEKGERLPQPEICTIDVYIMVCKMID 962
DB 897 SDVMSYVTVWELMTFGAKPYDGIAPREIPDLLEKGERLPQPEICTIDVYIMVCKMID 956
QY 963 SECPREPREVSEFRRARDPQRFVYIQ-NEDGSPSPDSTRYSLDDDDMGDIYDNE 102
DB 957 ADSRPKRELILLESKQARDPQRFVYIQDERMHLPSPTDSNFRALMDEEDMEDVYDAD 101
QY 1022 EYLVPOGFFCPDPAVGAGGVVHRRSSSTRSGGDLTLGLEPSEEARPSPLAPSEGA 108
DB 1017 EYLVPOGFFCPDPAVGAGGVVHRRSSSTRSGGDLTLGLEPSEEARPSPLAPSEGA 104
QY 1082 GSDVFGDGLMGAKGQSLPTDSEPLQRYSEDPVPLPSET--DGYVAPLTCSPQPEY 113
DB 1043 TSN---NSTVACINNGSCRYKEDAFLORYSDDPGAVTEDNIDDAFL-----FVPEY 109
QY 1140 VNQPDYAPCPSPREBPPLPAARAGATLERATLSPGKGVVQVDAFPGAVENPEYL-T 119
DB 1093 VNQ-SVPKRAGSVQNPVYHNOGLH-----ABGRLLHYQN-PSNAGNPEYLLT 114
QY 1199 PQGAAPQPPPPAPFPAFDNLVYMDQ-----DP-----PERGAPSTFKGTPT 124
DB 1142 AQ-----PCLSSGPNSPALMTQKSHQMSLDNPDYQDFFPKETKPKGIFKQ-PT 119
QY 1243 AENPEYIGLDPV 1254
DB 1192 AENAEYLRVAPP 1203

```

RESULT 6
ERB4 HUMAN
ID ERB4 HUMAN STANDARD; PRT; 1308 AA.
AC Q15303;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
DE (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
GN ERBB4 OR HER4 (Human).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM JM-A).
RC TISSUE=Breast carcinoma;
RX MEDLINE=93189574; PubMed=8383326;
RA Ploman G.D., Cuiouascu J.-M., Whitney G.S., Green J.M., Carlton G.W.,
Foy L., Neubauer M.G., Shoyab M.;
RT "Ligand-specific activation of HER4/p180erbB4, a fourth member of the
RT epidermal growth factor receptor family";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
RP SEQUENCE FROM N.A. (ISOFORM JM-B).
RC TISSUE=Fetal brain;
RX MEDLINE=97476287; PubMed=9334263;
RA Blenius K., Corfas G., Paul S., Choi C.J., Rio C., Ploman G.D.,
Klaegstrup M.;
RT "A novel juxtamembrane domain isoform of HER4/erbB4. Isoform-specific
RT tissue distribution and differential processing in response to
RT phorbol ester";
RL J. Biol. Chem. 272:26761-26768(1997).

CC - FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULIN. NR-
 CC 2, NR-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
 CC NTK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
 CC NOT ACTIVATED BY EGF, TGF- α , AND AMPHIREGULIN.
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC - SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
 CC RECEPTORS (POTENTIAL).
 CC - SUBCELLULAR LOCATION: Type I membrane protein.
 CC - ALTERNATIVE PRODUCTS: 2 ISOFORMS, JM-A (SHOWN HERE) AND JM-B;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
 CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
 CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
 CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
 CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
 CC - TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
 CC KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,
 CC PITUITARY, SPLEEN, TESTIS AND PANCREAS.
 CC - LUNG, SALIVARY GLAND, AND PANCREAS.
 CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES.
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: L07868; AAB59446.1; -
 DR HSSB: P11362; IFGK.
 DR Genew: HGNC:3432; ERBB4.
 DR MIM: 600543; -
 DR Interpro: IPR000494; EGFR_L_domain.
 DR Interpro: IPR000719; Euk_kinase.
 DR Interpro: IPR002174; Euk_kinase.
 DR Interpro: IPR001245; Tyr_kinase.
 DR Interpro: IPR004019; YLP_motif.
 DR Pfam: PF00069; Pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR Pfam: PF02757; YLP; 2.
 DR Prodom: P0000001; Euk_kinase; 1.
 DR SMART: SM00261; FU; 4.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN KINASE_TYR; 1.
 DR PROSITE: PS00111; PROTEIN KINASE_DOM; 1.
 DR Transmembrane: Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Alternative splicing.
 FT SIGNAL 1 25
 FT CHAIN 26 1308
 FT DOMAIN 26 651
 FT TRAMSMEM 652 675
 FT DOMAIN 676 1308
 FT DOMAIN 186 334
 FT DOMAIN 496 633
 FT DOMAIN 718 985
 FT NP_BIND 724 732
 FT BINDING 751 751
 FT ACT_SITE 843 843
 FT DISTLFD 189 197
 FT DISTLFD 193 205
 FT DISTLFD 213 221
 FT DISTLFD 217 229
 FT DISTLFD 230 238
 FT DISTLFD 234 246
 FT DISTLFD 249 258
 FT DISTLFD 262 289
 FT DISTLFD 293 304

FT DISTLFD 308 323 BY SIMILARITY.
 FT DISTLFD 326 330 BY SIMILARITY.
 FT DISTLFD 503 512 BY SIMILARITY.
 FT DISTLFD 507 520 BY SIMILARITY.
 FT DISTLFD 523 532 BY SIMILARITY.
 FT DISTLFD 536 552 BY SIMILARITY.
 FT DISTLFD 555 569 BY SIMILARITY.
 FT DISTLFD 559 577 BY SIMILARITY.
 FT DISTLFD 580 589 BY SIMILARITY.
 FT DISTLFD 593 614 BY SIMILARITY.
 FT DISTLFD 617 625 BY SIMILARITY.
 FT DISTLFD 621 633 BY SIMILARITY.
 FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 138 138 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 174 174 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 181 181 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 253 253 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 358 358 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 473 473 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 548 548 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 620 620 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VASPLIC 626 648 NGPSSHDCIYPTGCTSLPQHA -> IGSSIDCIGLMD (IN ISOFORM JM-B).
 SQ SEQUENCE 1308 AA; 146807 MW; 5E4BE095D8761 CRC64;
 Query Match 43.3%; Score 2950.5; DB 1; Length 1308;
 Best Local Similarity 45.0%; Pred. No. 4; se-150;
 Matches 606; Conservative 189; Mismatches 384; Indels 173; Gaps 28
 QY 9 WGLLLALLPFGAA-----STOVCTGTDMKRLPSPETHLDMRLHYQGCQVVOGNIETLY 64
 DB 8 WWSVSLVLAAGTVOPSDOSQVCACTERKLSLSLDELQYRAKRYENGEVWGNLEITS 67
 QY 65 LPTNASHLFDIDIOEVOQYIKANSKFIGITELRLIRVGTOLFEENVALVAVDNGDPLN 124
 DB 68 IEHRDLSFLRSVREYGVYVALNORFYLPLENLIRKTKYERVALAIFLNRKGG 127
 QY 125 NTPVTGASPGGLRELOLSLTELKGVLIQSNPOLCYQDTILMKDFHKNNQALLTLI 184
 DB 128 NF-----GLQELGLKNTLEILNGVVVDNKKELCYADTIHMODIYRNPVSNLTLV 178
 QY 185 DTNRSPACHPCSPWCKSRGCGSSSEDCSLTETVACGC-ARCKPRLPDCHEGCAAG 243
 DB 179 STNSSGCGRCHSCTG-KWGPTEHNCUTLITVCAEODGRCYGVYVSDCCRCAGG 237
 QY 244 CTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPBGRYTFGASCTACPYNY 303
 DB 238 CSGPKDIDFACMNFNDSGACVTCQPTFYVNTTQLEHNFPAKTYGAFCKKCPHF 297
 QY 304 LSTVGSCTIVCPPLHNGEYAEQGTQRCCKSPCARVCYGLGMEHLREVRATYNSIQR 363
 DB 298 V-VSSSCVAPASCKREV-EEGIRKCKPCTDICRACDGGTGLSMAQYVDSNDIK 355
 QY 364 FAGCKLFGSLAPFESFDGDPASNTAPLOPEOLQVFTLEETIGLYISAMPDLSLPS 423
 DB 356 FINTKXINGMLIFLVGSIHDPYNALEIDPELVNFRYREITGLNLSQNPNNMTDS 415
 QY 424 VFQVLQYIRGILHNGAVSLTQGLISGLSLELDSGLALHHNTHLCFVHTVPD 483
 DB 416 VFSNLVTIGRIVYSGSLILKQOQITSLQFOSLKEISAGNIYIINDSNLCYHTINTT 475
 QY 484 QLFNPNQALHTANRDEDCVEGGLACQLCARGCGWGPFGQCNGQPFRCGOE 543
 DB 476 TLFEITNORIVINDNRKAEVCTAEWVCHNLCSDDCKWGPBGQICSCRSRGRCTIS 535
 QY 544 CRVIQGLPREYVNAHRCCLPHEPCP-QNGSVTCFPEADQCVACHYDPFCVACRS 602

```

Db      536 CNIYDGEFFREFENGISICVECDPQCEKMEGDLTLTGHPGPDNCTCKSHFXDGNVCYKCPD 595
Qy      603 GVKPDLISYMIKMPPEBGACQCPINCHSCVDLDDKCC-----PAERASPL 651
Db      596 GLOGANSF--IFKYADPDRECHPCFNCQCGPTSHDCIYPTGSHSTLQOHAR-TPL 652
Qy      652 TSIVSAVY-GILVVVVLGVVFGILLKRRQOKIRKYMRLLOETELVEPLTPSGAMPNOA 710
Db      653 --IAAGVIGGLFLVIVGLTFAYVYARKSIK-KKRALRFL-ETELVEPLTPSGAPNOA 708
Qy      711 QMRILKTELARKKVLGSGAFGVYVYVGTIPDGENKIVALKVRENTSPANKELIDE 770
Db      709 QRLRLKTELARKKVLGSGAFGVYVYVGTIPDGENKIVALKVRENTSPANKELIDE 768
Qy      771 AYMAGVGSPPYVRLIGICTSTVQVLTOLMPGCLLDHVENRGLSGODLLNMCMOIA 830
Db      769 ALIMASMDHPLVRLILGVCLSPITQVLTOLMPHGLLEVHKNIGSOLLIMCVIOIA 828
Qy      831 KMSVLEEDVTLVRDLAARNVLKSNHYKITDFGLARLLIDETFEYHADGAKVPIKMA 890
Db      829 KGMVYLEERLTVHRDLAARNVLKSNHYKITDFGLARLLIDEGDEXYNADGKMPKMA 888
Qy      891 LESILRRRFTHSDVSYGYVWELMTFGAKPYDGI PAEIPDLLEKGERLPDPICTID 950
Db      889 LECIHRKFTHSDVSYGYVWELMTFGAKPYDGI PTREIPDLLEKGERLPDPICTID 948
Qy      951 VYMIYKCMWIDSECPREELVSEFSRMAPDPQRFVYIQND-LGPASPLDSTFYRSL 1009
Db      949 VYVYVYKCMWIDADSPPKFEELAEFSRMAPDPQRFVYIQND-LGPASPLDSTFYRSL 1008
Qy      1010 EDDMDGLVDAEYLVLPQCGFPQDPAAGGWMHHRSSSTRGCGDILLGLEPSEE 1069
Db      1009 DEEDLEDMDAEYLVLP-QAFNIPP-----LYSRAITDNRN-----ELGSPPAY 1056
Qy      1070 APRS-----PLAV-SEGAQSDVAFDGLDGLGAKGLQS 1100
Db      1057 TPMGNGQFYVRCGFAAEQGVSVYRAPSTIPBAVAGAAELFDSCNGTIRKVA 1116
Qy      1101 LPTHDSPLQVSEDTVPPLS-----ETNGYVAPLTCSPQRYVQDPVRFQPPSPR 1153
Db      1117 PHVQEDSSTQRYSDADTVAPERSPERGELDEGYMTPMKDKCKOELNPNVE----- 1167
Qy      1154 EGPLPAPAPAGATLEBAKTLSPKNGVVDVAFGAVENPEYLPQGGAAPOHPPPA- 1212
Db      1168 ENPFVSR-----KQNDLQ-----ADNPEYHANSNG-----PPAAE 1199
Qy      1213 -----FSPAPDNLVYWDQDPPEGA--PPESTF 1237
Db      1200 DEYVNEPLYLNTFANTLGRAEYLNKNNILSMPEKAKKAFEDNPDYMHSLPPESTLQHPDYL 1259
Qy      1238 KGTPT-----AENPEYL 1249
Db      1260 QEYSTKYFYKQNGRIRPIVAENPEYL 1285

```

```

RA      Marchionni M.A., Kelly R.A.;
RT      "Neuregulin promotes survival and growth of cardiac myocytes.
RT      Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT      ventricular myocytes."
RL      J. Biol. Chem. 273:10261-10269(1998).
RN      [2]
RP      SEQUENCE OF 848-901 FROM N.A.
RC      TISSUE=Sciatic nerve;
RX      MEDLINE=9122560; PubMed=2025425;
RA      Lai C., Lemke G.;
RT      "An extended family of protein-tyrosine kinase genes differentially
RT      expressed in the vertebrate nervous system";
RL      Neuron 6:691-704(1991).
RN      [3]
RP      SEQUENCE OF 1031-1196 FROM N.A.
RC      STRAIN=Sprague-Dawley; TISSUE=spinal cord;
RX      MEDLINE=9718422; PubMed=9030624;
RA      Carroll S.L., Miller M.L., Frommert P.W., Kim S.S., Corbett J.A.;
RT      "Expression of neuregulins and their putative receptors, ErbB2 and
RT      ErbB3, is induced during Wallerian degeneration.";
RL      J. Neurosci. 17:1642-1659(1997).
CC      -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULIN, NRG-
CC      2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC      NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC      NOT ACTIVATED BY EGF, TGF- $\alpha$ , AND AMPHIREGULIN (BY SIMILARITY).
CC      -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC      RECEPTORS (POTENTIAL).
CC      -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC      NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC      OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC      RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC      HEART.
CC      -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC      RESIDUES (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      EMBL: AF041838; AAD08899.1;
CC      EMBL: U52531; AAC53051.1;
CC      HSSP: P11362; IREGK.
CC      DR InterPro: IPR000494; EGFR_L_domain.
CC      DR InterPro: IPR000719; Euk_pkinase.
CC      DR InterPro: IPR002174; Furin-like.
CC      DR InterPro: IPR001245; Tyr_pkinase.
CC      DR InterPro: IPR004019; YLP_motif.
CC      DR Pfam: PF00757; Furin-like; 1.
CC      DR Pfam: PF00069; pkinase; 1.
CC      DR Pfam: PF01030; Recep_L_domain; 2.
CC      DR Pfam: PF02757; YLP; 2.
CC      DR PRINTS: PR00109; TYRKINASE.
CC      DR PRODOM: PDD00001; Euk_pkinase; 1.
CC      DR SMART: SM00261; FU; 4.
CC      DR SMART: SM00219; TYRKC; 1.
CC      DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC      DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
CC      DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
CC      DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC      FM transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
CC      FT SIGNAL 1 25 POTENTIAL
CC      FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
CC      FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
CC      FT TRANSMEM 652 675 POTENTIAL.
CC      FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).

```



```

FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 CYS-RICH.
FT NP_BIND 718 985 ATP (BY SIMILARITY).
FT BINDING 724 732 ATP (BY SIMILARITY).
FT ACT_SITE 751 751 ATP (BY SIMILARITY).
FT DISULFID 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 PYT -> SYR (IN REF. 3).
SQ SEQUENCE 1308 AA; 146957 MW; D944BB0996A08B41 CMC64;

```

Query Match 43.0%; Score 2931; DB 1; Length 1308;
 Best Local Similarity 44.8%; Pred. No. 4.9e-149;
 Matches 604; Conservative 190; Mismatches 393; Indels 160; Gaps 28;

```

QY 1 MELA-ALCRMGILT--ALLPFGAASQTGCTGDMKRLRIPASPELTHDMLRHLYOGGVQV 57
DB 1 MKLATGLMWGSLVLAARTVOPSHASQSVCACTENKSLSDLEQYRALRKYYENCEVVM 60
QY 58 GNLELYLPTNASLSTFQDIQEVQYIKANSKFTIGTELQRLRIVRGTOIFEDNYALVL 117
DB 61 GNLEITSIENHRDLSFIRSIREVTVGLVALNQFRYLPLENLRIIGTKLYEDRYALAF 120
QY 118 DNGDPLNNTTPVTCASGGLRELDLRSLTEILKGVYLICQNPOLCYQDITLIMKDIFFHKN 177
DB 121 LNYRKQGNF-----GLOELGLKNLTLEILNGGVVDQNKFLCYADITTHQDIVRNF 171
QY 178 QALATLIDTNRSRACHPCSPWCKSGRCWGESSEDCSLTRTVACAGGC-ARCKGFLPTDCC 236
DB 172 PSNMTLVSTIGSSCGRCHKSCG-RCMGPTENHCQILTRTVCAEQCDGRKGVYSDCC 230
QY 237 HECCAAGCTGPKHSDCLAHFNHSGICELHCPALVTYNTDTRESNPNEGRTYTRASCY 296
DB 231 HRCAGGSGPKDTCACAKMFPNDSGACVTQCPOTFVYNPFTFQLEHNNAKTYGAFV 290
QY 297 TACPYNLTSDVSGCTLVCPDLHQEVTAEDGTQRCCKSKPCARVCGYGLMEHLREVAV 356
DB 291 KKCCHNFV-VDSSSCVNACPSKMEV-EENGIMCKRCPTDICAKADGIGTGSLSMSAQTV 348

```

```

QY 357 TSANIOEFACCKXIFGSLATPESFPDGPASNPALQPEQLQVFTLEETIGLYISAMP 416
DB 349 DSSNIDFNFICTKINGLLIFLVTHGHDPNNAIDALDPKLNFRVTRITGLNLQTP 408
QY 417 DSLPDLSPFQNLQVIRGRILHNGAVSLTQGLGISVLGHSRLRELSGGLALIHNTLFC 476
DB 409 PNMIDFVFNLVTIGGRVLYSGLSLLILKQGITSLQOSLKEISAGNIYITDNGNLV 468
QY 477 VHTVPMQLEFRNPHQALLHTANRPEDECYGBGLAQHLCARHGCHGPGPTQCVCNCSQFLR 536
DB 469 YHTIMNTLTFSYVQRVIRIDNRANCTAEGVNCNHLCSNDGCPGPDQCLSCRRFSR 528
QY 537 GOECVEECRLQGLPREYVVARHCLPCHPECCP-ONGSVTCFPEZDQCVACAHYDPDF 595
DB 529 GKICIECNLYDEFEFENGSLCVCEBQCEKMEGDLTTCGPRGDNCTKSHFMDGN 588
QY 596 CVARCPGVKPDLSYMPWKPFDEBACOPCPINCTHSCVLDLDDKQC-----PA 644
DB 589 CVEKCPVOLGANSF--IFKYADQDRCHCHENCTQCGNGPFSHDCIYVPTGSHSLPQ 646
QY 645 EQRASPLTISVAVV-GILLVVLGVVFGILIKRQOKIRKYMRLLOETLEVEPLTFS 703
DB 647 HAR-TPL--IAAGVIGSLVLTALTPAVYARKSK-KKQALRRFL-ETLEVEPLTFS 701
QY 704 GAMPNQAQMRILKETEELRYKVLGSAFGTVYKGIWIPGENVKIPVAIKVLRNTSPXA 763
DB 702 GTANQQAQMLILKETELRYKVLGSAFGTVYKGIWIPGENVKIPVAIKVLRNTSPXA 761
QY 764 NKEILDAAVMAVGSGSYVSRLLIGLITSLVQLVLTQMLMYGCLLDHYRNRGRLSGDL 823
DB 762 NVEWMDALMAVDHVLVRLGLVCLSLPTQLVLTQMLMYGCLLDHYRNRGRLSGDL 821
QY 824 NMCGQIAKMSYLEDVRLVYRDLAANVLYKSNHXYITDFGLARLLDDETEYHADGK 883
DB 822 NMCGQIAKMSYLEDVRLVYRDLAANVLYKSNHXYITDFGLARLLDDETEYHADGK 881
QY 884 VPIWMALESILRRFTQSDVSYGVYTWELMTFGAKYVDGIIPAREIDLLKGRFLQ 943
DB 882 MPIMWMALESILRRFTQSDVSYGVYTWELMTFGAKYVDGIIPAREIDLLKGRFLQ 941
QY 944 PPICTIDVYIMVCKMWDSECPREELVSEFSRMAPRQPRVVIQNEB-LGPA SPLDS 100
DB 942 PPICTIDVYIMVCKMWDSECPREELVSEFSRMAPRQPRVVIQNEB-LGPA SPLDS 100
QY 1003 TFYRSLDEDDMDGLVDAEYLVPOQGFCCPDP-----APGA 103
DB 1002 KFFQNLDEDDMDGLVDAEYLVV-QAFNIPPIYTSRTRIDNRSEIGHSPPAYTPMS 106
QY 1040 GGVVHHNRSSSTRSGGDLTLGLESEEBAPRSLAPSEGAQSVDFDGLMGAAKGLQ 109
DB 1061 GSGVYVODGSPATQCG--MPMPTATTTSTIPAPVA--QGATAENFDDSCCNGLTRKFV 111
QY 1100 SLPTHDSPLOQRYSEDPVLPFS-----ETDGYAPLTCSPQEPYVQAPVRPQPSF 115
DB 1116 VPHVQDESSSTRQYSADPTVPAPERNRAELDEGVTMPHMDKPKQVILPVE----- 116
QY 1153 REGPLPAPRAGATLEBAKTLSCGKNGVYKDVAFAGAVENBEYLTLPQGAAPQHPRA 121
DB 1168 -ENFVFSRR-----KXGDLQ-----ADNDEYHSASSG-----PPRA 119
QY 1213 -----FSPAFDNIYYMQDPPERGA--PST 123
DB 1199 EDEVNRPVLTNFTNALGNAEYMKSSLSVPEKAKKADNDYMNHSLPPRSTLQHPY 125
QY 1237 FKGTPT-----AENPEYL 1249
DB 1259 LOEYSTYFYKQNGRIRPIVAENBEYL 1285

```

RESULT 8
 XMRK_XIPMA
 ID XMRK_XIPMA STANDARD; PRT; 1167 AA.
 AC P13388;


```

Db      591 PRCHGILGSDTL-1WKVADKMGOCOPCHQNTQCCSGPGLSGCCGD-IVSHSLAIVL 648
Qy      658 VGIILVVLGVVFGIILIKRQOKIRKRYMRLLQETLVEPLTPSGAMPNQQRILKE 717
Db      649 VSGILITVIVALLIVLIRRRIRK-RKRITRCLQCKELVEPLTPSGQAPNQAFIRLKE 707
Qy      718 TELRKVAVLGSAGVYKGIWIPDGENYKIPALIVAEINSPRANKELIDEAVYAGV 777
Db      708 TERKQDVLSGSAHGVYKGLAMPDSGNIRIPALIVAEINSPRANKVQVLEAVYASV 767
Qy      778 GSPYVSRLLGICLTSTVQVLTQMPYGCCLDHYENRGLSGODLLNMQOIAKMSYLE 837
Db      768 DHPHVCRLLGICLTSAVQVLTQMPYGCCLDHYRQGERICGQWILNVCQIAKMYNLE 827
Qy      838 DVLVHRDLAARVAVLSPHNVKITDGLARLLDIDETRYHADGKVPYKMALESILR 897
Db      828 ERHLVHRDLAARVAVLSPHNVKITDGLARLLDIDETRYHADGKVPYKMALESILR 887
Qy      898 RFTQSDVMSVGVTVWELMTFGAKPYDGIAPAREIPDLLEKGERLPQPICTIDYVMIVK 957
Db      888 TYTHQSDVMSVGVTVWELMTFGSKPYDGIAPAREIPDLLEKGERLPQPICTIDYVMIVK 947
Qy      958 CMWIDSECRPFRELVSESRVARDPQRFVYVIONEDLGPASPLDSTFYRSLIEDDMDL 1017
Db      948 CMWIDSECRPFRELVSESRVARDPQRFVYVIONEDLGPASPLDSTFYRSLIEDDMDL 1002
Qy      1018 VDAEEYLVPOGFPFPPAPAGAGVWHRRSSSTRSGGDDLGLGLEPEEEAPRSPAP 1077
Db      1003 VDAEYLLPYKRT-----NRQGS-----EPCLP 1025
Qy      1078 SEGAGSDVFDGLGMAKAGLSLPTHDPSPLOVSEDEPTV-PLPSETDGYVAPLTCSPQ 1136
Db      1026 PTGH-----PYRENSITLRNISDPTQVNALEKDLGH----- 1056
Qy      1137 PEYVNOQDVAPQP-----PSPRE-----GAP-ARARAGATLEKAKLSGKGVYKD 1183
Db      1057 -EYVNOQGETSSRLSDIYNPNEDLTDGMPVSLSSQAEITNFSPEYLLNNQSL--- 1112
Qy      1184 VFAFGAVENPEYLLTPGCGAAPPHPPAFSPAFNDLYWDDPPERGAPESTFGCTPTA 1243
Db      1113 PLVSSGSMDDPDY---QAG-----YQAAF-----LPQTGALLTNGMFLDPA 1150
Qy      1244 ENPEYLG 1250
Db      1151 ENLEYIG 1157

RESULT 9
ERB3_HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
CT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (BC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90063234; PubMed=2687875;
RA Klaus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors."
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RL [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todor G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene."
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7665162;
RA Kato M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase."
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULIN AND NTAK
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAIN OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE p45 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M29366; AAA35790.1; -
DR EMBL; M34309; AAA35978.1; -
DR EMBL; S61953; AAB26935.1; -
DR F1R; A36223; A36223.
DR HSP; P11362; 1FGK.
DR Genew; HGNC:3431; ERBB3.
DR MIM; 190151; -.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_Kinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Prodom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FTY_3.
DR SMART; SM00219; TYKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Alternative splicing
FT SIGNAL 1..19
FT CHAIN 20..1342
FT DOMAIN 644..664
FT TRANSSEM 644..664
FT DOMAIN 665..1342
FT DOMAIN 709..966
FT NP_BIND 715..723
FT BINDING 742..742
FT ACT_SITE 834..834
FT DISULFID 186..194
FT DISULFID 190..202

```

```

FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.
FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 266 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 552 565 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 141 183 EILSGVYIEKNDKLCHEMDITDWRDIDRDAIVKDNDR
SC -> GQPMVPSGLPPQADQWYLLDDPRLTLTASASK
VPTLAIV (IN SHORT ISOFORM).
FT VARSPLIC 184 1342 MISSING (IN SHORT ISOFORM).
FT CONFLICT 560 560 E -> G (IN REF. 2).
FT CONFLICT 1064 1064 E -> G (IN REF. 2).
SO SEQUENCE 1342 AA: 148097 MM: 7201E7F66CA374BD CRC64:

```

Query Match 35.2%; Score 2398.5; DB 1; Length 1342;
 Best Local Similarity 40.3%; Pred. No. 1.3e-120;
 Matches 528; Conservative 189; Mismatches 465; Indels 129; Gaps 32;

```

QY 10 GLLALLPPGA--STOVCTGTDMKRLPASPTHLDMLHLYOGQOVQGNLETTYPT 67
DB 11 GLFLSLARGSEVNSQAVCPGTLNGLSVTGDAENQVOTLYKLYERCEVVMGNLEIVLGH 70
QY 68 NASLFLQDIOEQVYIKANSKFIQITELQRLRIVRGTOLEFENVALAVLDNGDPLNNT 127
DB 71 NADSLFQWIREVYGVLVANNESTPLPLNLYAVGTQVYDKKALFVM-----LNIYT 125
QY 128 PVTGASPGGLRELOASLTLEILKGVLIQNNPOLCYQDTILMKDIFHKNOALATLIDTN 187
DB 126 ----NSSHARLRLTLQTLTEILSGVYIEKNDKLCHEMDITDWRDIDRDAIVKDNDR 178
QY 188 RSRACHPCSPMKCKSRGWSSEDCOSLTITVCAAGC-ARCKGRLTDDCHCQCAAGCTG 246
DB 179 NGRSCPCPCHEVCKG-KCWGPGSEDCOTLTKTICAPQCNHCFEPNNQCHCECGGCGG 237
QY 247 PKGSDCLALHFNHSGICELHCPALVTYNTDFESNPNEGRTFGASCVTACPNYVLTST 306
DB 238 PDDTDCFACHFNDSGACVPCRPQPLVYNKLTQLEPNHTKXQYGVGVASCPHNFV-V 296
QY 307 DVGSCITVCPHNLQEVTAEDGTQRCCKSPKACVYVGIQMEHLREAVTASANTQEPAG 366
DB 297 DQTSVCVACPPDKNEVD-KNGLMKCPGGLCPKACGEGGSG--SRFQTVDSNSNDGFVN 353
QY 367 CKKIFGSLAEFLPESFDQDPASNTAPLOPELOVFEFLTEITGYLYSAPDLSPLSVFQ 426
DB 354 CTKILGNLDFPLIQLNGDDPWHKIPALDPEKLVNFRVREITGYLVNQSPPHMHNFVSFS 413
QY 427 NLQVIRGRILHNAGYS-LTIGGIGTSMGLKRSIRELGSGLALIHNTLTCFVYTVPMQDL 485
DB 414 NLTTIGGRSLYNGFSLIMKNLNVTSIGFRSLKETISAGRIVISANROLCHYHNSLMTKV 473

```

```

QY 486 FRNPQALLHTA-NRPEDECVGEGLACHOLCARGHCGWPGPTQVCNCSQFLRQCEVEEC 544
DB 474 LRGPTEERLDIKHNRPRDCAVEGKCDPLCSSGGCGPQGLSCBRYNSRGVCVTHC 533
QY 545 RVLOGLPREYVNAHNCIPCHPECOQPNQSVTCTGPEADQCAAHKKDPFCVACRPSGV 604
DB 534 NPLNGEPREFAHEAECSCCHPECOQPMGTATCNGSGSDTCAQCAHFRDGHCVSSPHGV 593
QY 605 KPDSLVPWIKFPEDEACQPCPINTCTSCVDLDDKCPAEGRA-----SPLTSIVAAVVG 660
DB 594 LG--AKGIYKVPVQNECRCHENCCTGCKGSELDDCGLTLVLGKHTLNTALVIAG 651
QY 661 ILLVVLGVVFGILIKRQOKIR-KYTMRELLQETELVEPRTSGAMPQAOQRILKETE 719
DB 652 --LVVIFPMQLGFTLYWRGRRIQNKRAMREYTERGESIEPLDPS-EKANVYLARIFEKE 708
QY 720 LRKXVLSGAGFTVYVGIWIPQSENVKIVAIKVLRENTSPRANKEILDEAVVAGVGS 775
DB 709 LRKXVLSGAGFTVYVGIWIPQSENVKIVAIKVLRENTSPRANKEILDEAVVAGVGS 768
QY 780 PYVARLLGICLTSTVQLVLTQJLMPYGLLDHVRNNGRLSGQDLNNCMQIAKQMSYLEDV 839
DB 769 AHIVRLGLCGSSGLQVLTQYLPFGSLDHRQHRGALGQULLNNGVOIAKQMYLIEBH 828
QY 840 RLYVRDLAARVLYKSPNHYKITDFGLARLLDDETYHADGKVPYIKMALESILRRRP 899
DB 829 GMYHRNLAAARVLLKSPQOVAVDFGADLLPDDQQLYSEAKTPIKMALESISHFGX 888
QY 900 THQSDVMSYGVYVWELMTFGAKPYDGIIPAREIDPLEKEGRLEPOPCITIDVYMIWKCW 959
DB 889 THQSDVMSYGVYVWELMTFGAEFYAGIRLAIEVDLLEKGRLEPOPCITIDVYMIWKCW 948
QY 960 MIDSECRPRELVESEFSRNARDPQRFVVIQNEDLGPA---SPLDSTFYRSLLEDDMGD 101
DB 949 MIDENIRPTEKEALNEBTPMARDPRLVIVIKRES-GPGIAPGSEPHGLNKKULEVELEP 100
QY 1017 LVDAEYLIVQOQGFCDPAPAGAGWVHHRHSSSTRSGDGLTLELPE-SEEAARSEL 107
DB 1008 ELDDLLELEAED-----NLATTLGSAALSLVGTUNRPRGQSLL 104
QY 1076 APSEGAGSDYFDGDLGAKAGLQSLPTH-PSPLQVSEDPFVPLP-----SEFDGVY 112
DB 1049 SPSSGQ-MPMNQNLGSSCGESAVSSSECRPPVSLH-----PMRQCLASSESGHY 110
QY 1129 A-----PLTCSQPF-----YVNPQPVVRQPPSPRECP-----L 115
DB 1102 TGSBAELQEKVSMCRSRSRSPRPGDSAYHNSQHSLLTPVPLSPGLEBEDVNGVYM 116
QY 1158 PAARFAGATLERAKTISP-GKNGVV-----KQVFAFGAVENPEYLTPOGGAAPQPHP 121
DB 1162 PDLTKCTPSSREGTILSVGLSSVLTGEEDD-----EVEYIMNRRRHNSP-PHP 121
QY 1211 PAFSPAFNLYVD-----QDPERCAPSTFKGTPTAENPEYL 1249
DB 1213 RPSLELELGEYVMDVGSLSLSTQSCPLHPVIMPPTAGTTDPDEDVYM 1263

```

RESULT 10
 ERB3_RAT
 ID ERB3_RAT STANDARD; PRT; 1339 AA.
 AC Q62799; Q62955;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor Protein-tyrosine kinase etbB-3 precursor (EC 2.7.1.112)
 DE (C=etbB3).
 GN ERB3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RA MEDLINE=96096535; PubMed=8522190;
 RX Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
 RT "Cloning of the rat ErbB3 cDNA and characterization of the
 RL recombinant protein.";
 RN Gene 165:279-284(1995).
 RP REVISIONS TO 85; 513 AND 565.
 RA Hellyer N.J., Koland J.G.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RP [3]
 RP SEQUENCE OF 922-1097 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
 RX MEDLINE=97184212; PubMed=9030624;
 RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
 RT "Expression of neurotrophins and their putative receptors, ErbB2 and
 RL ErbB3, is induced during Wallerian degeneration.";
 RJ J. Neurosci. 17:1642-1659(1997).
 CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTRK.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
 CC PHOSPHATIDYLINOSITOL 3-KINASE.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; U29339; AAC28498.2; -;
 DR EMBL; U52530; AAC53050.1; -;
 DR HSSP; P11362; 1FGK.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR000719; Furin_kinase.
 DR InterPro; IPR002174; Furin-like
 DR InterPro; IPR001245; Tyr_kinase.
 DR Pfam; PF00069; Kinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR PRINTS; PRO0109; TYRKINASE.
 DR ProDom; PD000001; Euk_kinase; 1.
 DR SMART; SM00261; FU; 5.
 DR SMART; SM00219; TyKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR Trnsmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 19
 FT CHAIN 20 1339
 FT DOMAIN 20 643
 FT TRANSMEM 644 662
 FT DOMAIN 663 1339
 FT DOMAIN 183 259
 FT DOMAIN 707 964
 FT NP_BIND 713 721
 FT BINDING 740 740
 FT ACT_SITE 832 832
 FT DISULFID 186 194
 FT DISULFID 190 202
 FT DISULFID 210 218
 FT DISULFID 214 226
 FT DISULFID 227 235

FT DISULFID 231 243 BY SIMILARITY.
 FT DISULFID 246 255 BY SIMILARITY.
 FT DISULFID 259 286 BY SIMILARITY.
 FT DISULFID 290 301 BY SIMILARITY.
 FT DISULFID 303 320 BY SIMILARITY.
 FT DISULFID 323 327 BY SIMILARITY.
 FT DISULFID 500 509 BY SIMILARITY.
 FT DISULFID 504 517 BY SIMILARITY.
 FT DISULFID 520 529 BY SIMILARITY.
 FT DISULFID 533 549 BY SIMILARITY.
 FT DISULFID 556 573 BY SIMILARITY.
 FT DISULFID 576 585 BY SIMILARITY.
 FT DISULFID 589 610 BY SIMILARITY.
 FT DISULFID 613 621 BY SIMILARITY.
 FT DISULFID 617 629 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 353 353 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 408 408 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 414 414 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 437 437 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 469 469 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 522 522 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 566 566 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 616 616 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CONFLICT 1028 1028 L -> P (IN REF. 3).
 SQ SEQUENCE 1039 AA; 147545 MW; 0A5F2402BDFD1E CRC64;
 Query Match 34.3%; Score 2332.5; DB 1; Length 1339;
 Best Local Similarity 40.5%; Pred. No. 4.2e-117;
 Matches 520; Conservative 170; Mismatches 438; Indels 155; Gaps 34
 QY 3 LAALCRGGLLALLPGAA---STQVCTGDMRLRASPETHLDMRLHYOCQVYQGN 59
 DB 7 LQVLC---FLSLARGSEWNGNQAVCPGLNLSTGADNGYQTLTYLKECEVWNGN 62
 QY 60 LELTYLPTNALSIFLDIQEVQYITANSKFITGLQLRYRGTLQFEDNYALAVLDN 119
 DB 63 LEIVLGNHNDLSFLQVIREVTGYVWANNESVLPPLMLRVRRGQVYDGFALFVM-- 120
 QY 120 GDPPLNTTPVTGASPGSLRELQRLSTELIKGVALIQRNPOLCYOPTILMKDIFHNGQL 179
 DB 121 ---LNNVT---NSSHALRQLKTLQLEITLISGVYLRKDKLCHMTIMRDLVR--- 170
 QY 180 ALTIIDNRSRAPHGSPGCKSGRCKGSESEDCQSITRVCAAGC-ARCKGPLPTDCHE 238
 DB 171 GAELIVKNGANGCPGCHVEYCKG-RWGPGRDDQILTKITICAPQCGRCFGPNQCCHD 229
 QY 239 QCAAGCTGPKHSDCLCLHFNHSGICELHCPALVTYNTDTPESMPYRPGRYFGASCVTA 298
 DB 230 ECAGGCGSPQDTDFACRFKRNDSGACVPRCPPELVYNNKLTFOLEPNPHTKYQGVCVAS 289
 QY 299 CPYNYISTDVSGCTLVCPPLHNOEVTAEQTORCEKSKPCARVYCYGLGMEHLREYAVTS 358
 DB 290 CPNHFV-VDQTFPCVACPPDKMEVD-KHGLKMEPCGGLCPACBEGTSGC--SRVQTVDS 345
 QY 359 ANIQEFPGCKIKIGSLAFLPESDGPASNTAPLOEBOUVEETLEITGYLYISAMPDS 418
 DB 346 SNIDGFNCKIKIIGNDLFTGLTGLNVPMKIPALDEKLVNRTVREITGYLYINISWEPH 405
 QY 419 LPDLVFNQNLVTRGRLIHNGAYS-LTLOGLGISWLGSLRELISGGLLIHNNHLCFV 477
 DB 406 MHNVSFVSNLTITIGRSLVYRGSLLIMKNLVNTSLGFSRLKISAGRYVIAANOQLCH 465
 QY 478 HTYPMWDLFPNPHQALIHRA-NPEDECYEGLAGHQLCARHGCWPGFTQCVCNSQFLR 536
 DB 466 HSLNMTRLTGPSEERLDIKYRFLDECLAEKGVCDPLCSGCGCWGPGGQCLSRNYSR 525
 QY 537 GQECVEECRVTLQGLPREYVNAHRLCPHPECOPONGSVTCFPEADQCYACAHYDPPFC 596
 DB 526 EGVGVTHCNLTQGERPFPVHEADQCSCHCECLPMESTSTCNSSGSDACRCAHFDGSHC 585
 QY 597 VAPCGVAKFDLSYPMIKWFPDEBAGACQPCPINCTHSC--VDLDDKCGCPAEGRASPLTSI 654

```

Db      586 VNSCPHGIIG--AKGIYKYPAONECEPCHECHCTOGCGNPELODCLQAQAVLMSKPHLV
Qy      655 VSAVGLILVWVIGVFGILIKRPOOKR-KYTMRLIOETLVEPLPSGAMPOQAQR 713
Db      644 IAVTVG--LAVILMILGSGFLVWRGRIONKAMRYLERGESIFLPDS-EKAKVLAR 700
Qy      714 ILKETELRKVYVIGSAGFTVYKGIWIPDGENVKI PVAIKVLENTSPKANKETIDEAYV 773
Db      701 IFKETELRLKLVIGSGVFGTVHKGIMIPGESIKIPVCIKVIEDKSGRQSPQAVTDHMLA 760
Qy      774 MAGVGSPPVSRLLIGITLTQVLTQVLMVYGLDHWENGRGSGQLLWMCQIAGKM 833
Db      761 VGSLDHAAHIVRLGLCPGSSLDQVLTQVLEGLDHWQHEHTGPPQLLWVGQIAGKM 820
Qy      834 SYLEDVRLVHRDLAARNVLYKSPNHVKITDFFGLARLIDIDETEVHADGKVPKIMALES 893
Db      821 YLLEHSMVHRDLAARNVLYKSPQVAVDFGVADLLPPDQQLHSEAKPIKIMALES 880
Qy      894 ILRRRTHSGDVSVGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPPOPICTIIVYM 953
Db      881 IHGKTYHOSDVWSGVTVWELMTGAPYAGLRALAIPLLEKGERLAPOICTIIVYM 940
Qy      954 IMVKCMWIDSECRPFRELVSEFSRMADPOGFVVICNEDLGPASPLDSTFYRLIEDD 1013
Db      941 VMVKCMWIDENIRPTFKELANFTMARDPRIYVITKAS-GPGTP--PAAEPSVLTATKE 997
Qy      1014 MGDVDAEYVLPQCGFCPPDPAAGAGVHHRRSSRSGGSDLTGLPSEEE----- 1068
Db      998 L-----QEAELEPEL-----DLIDLEAELEGATLS 1023
Qy      1069 -----EAPRPLAPSEG-----AGSDVEGDLGMAAGLOSULTHD 1105
Db      1024 LGSALSLPTGTLTRRGSOSLSPSSGMPWQNSLGACLDASVLRGREFSPISLH- 1082
Qy      1106 PSPLORYSEDPVTPVLPSETDGV-----APL-----TC-----SPOPE-----YVNPQDV 1145
Db      1083 PIPGR-----PASESEGHVTGSEAELOEKVSVCRSRSPRPROGSAVHSRHS 1135
Qy      1146 RQPPSPREGP-----LPARAGATLEAKTLP-GKQGV-----KDYFAF 1187
Db      1136 LITVTPUSPPELEEDGNGVMPPTHURGASSRGLTSSVGLSVLGTDEED----- 1191
Qy      1188 GAVENPEYLPQGAAPQPHPP 1210
Db      1192 -----EEVEYMARKRGRSP-FRRP 1209

RESULT 11
EGFR_DROME
ID EGFR_DROME STANDARD; PRT; 1426 AA.
AC P04412; O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05. Created)
DT 15-DEC-1998 (Rel. 37. Last sequence update)
DT 16-OCT-2001 (Rel. 40. Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE (Guinea receptor) (torpedo protein) (Drosophila relative of ERBB)
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schubach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in
RT subdomains of the receptor protein.";
RL Genetics 137:531-550(1994).
RN [2]
RP REVISIONS.

RA Clifford R., Schubach T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=45124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains.";
RL Cell 40:599-607(1985).
RN [4]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RX STRAIN=Oregon-R; TISSUE=Embryo;
RC MEDLINE=87002474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcript.";
RL Cell 46:1091-1101(1986).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RT specification in wild-type, Ellipse, and null mutant Drosophila.";
RL Dev. Biol. 205:129-144(1999).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Horkins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.C., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Boltskov S.,
RA Borrova D., Botchan M.R., Bouck P., Brockstein P., Brotlier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Deicher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferriz C., Ferriter S., Fleischmann W.,
RA Fioder C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasner K.,
RA Glisler A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jatala M., Kainush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paclet J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reibert K., Remington K., Saunders R.D.C., Schejter F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Splitter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.C., Wu D., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong P.N., Zhong W., Zhou X., Zhu S., Zhu X.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [7]
RP SEQUENCE OF 959-1078 FROM N.A.
RX STRAIN=Daekwanryeong;
RC MEDLINE=85137938; PubMed=2983232;
RA Wadsworth S.C., Vincent W.S. III, Bilo-deau-Wentworth D.;
RT "A Drosophila genomic sequence with homology to human epidermal

```

RT growth factor receptor";
 RL Nature 314:178-180(1985).
 RN [8]
 RP SEQUENCE OF 1133-1137, 1155-1157 AND 1209-1216 FROM N.A., AND MUTATION
 RP ANALYSIS.
 RX MEDLINE=92038942; PubMed1936959;
 RA Raz E., Schjeter E.D., Shilo B.Z.;
 RT "interallelic complementation among DER/flb alleles: implications for
 RT the mechanism of signal transduction by receptor-tyrosine kinases";
 RL Genetics 129:191-201(1991).
 RN [9]
 RP REVIEW.
 RX MEDLINE=97248481; PubMed=9094709;
 RA Perrimon N., Perkins L.A.;
 RT "there must be 50 ways to rule the signal: the case of the Drosophila
 RT EGF receptor".
 RL Cell 89:13-16(1997).
 CC -1- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
 CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
 CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
 CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
 CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
 CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOSEROSA
 CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
 CC CUTICLE.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
 CC PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND
 CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: UBICUOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
 CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANTAGON OF
 CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW. LEVELS REMAIN HIGH
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
 CC AND THORACIC AND ABDOMINAL GANGLIA.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF052754; AAC08536.1; JOINED.
 DR EMBL; AF052753; AAC08536.1; JOINED.
 DR EMBL; AF052754; AAC08535.1; JOINED.
 DR EMBL; AF052752; AAC08535.1; JOINED.
 DR EMBL; K03054; AAAS1462.1; JOINED.
 DR EMBL; K03417; AAAS1460.1; JOINED.
 DR EMBL; K03416; AAAS0965.1; JOINED.
 DR EMBL; K03418; AAAS1461.1; JOINED.
 DR EMBL; AF109077; AAD26134.1; JOINED.
 DR EMBL; AF109078; AAD26132.1; JOINED.
 DR EMBL; AF109082; AAD26132.1; JOINED.
 DR EMBL; AF109078; AAD26133.1; JOINED.
 DR EMBL; AF109084; AAD26133.1; JOINED.
 DR EMBL; AF109079; AAD26130.1; JOINED.
 DR EMBL; AF109081; AAD26130.1; JOINED.
 DR EMBL; AF109079; AAD26131.1; JOINED.
 DR EMBL; AF109083; AAD26131.1; JOINED.
 DR EMBL; AF109080; AAD26135.1; JOINED.
 DR EMBL; AE003454; AA46732.1; JOINED.
 DR EMBL; X02293; CA426157.1; JOINED.
 DR EMBL; X78920; CA455523.1; JOINED.
 DR EMBL; X78918; CA455522.1; JOINED.

DR EMBL; X78919; CA455522.1; JOINED.
 DR PIR; A00640; GQFE.
 DR HSSP; P11362; IFBK.
 DR Flybase; FBgn0003731; Egfr.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_Pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_Pkinase.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_Pkinase; 1.
 DR SMART; SM00261; FU; 7.
 DR SMART; SM00219; TYRK; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 DR TrEMBL; P55001; PROTEIN_KINASE_DOM; 1.
 KW Tyrosine-protein kinase, ATP-binding, signal, alternative splicing;
 KW Developmental protein.
 FT SIGNAL 1 30 POTENTIAL.
 FT CHAIN 31 1426 EPIDERMAL GROWTH FACTOR RECEPTOR.
 FT DOMAIN 31 868 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 869 889 POTENTIAL.
 FT DOMAIN 890 1426 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 938 1198 PROTEIN KINASE.
 FT NP_BIND 944 952 ATP (BY SIMILARITY).
 FT BINDING 971 971 ATP (BY SIMILARITY).
 FT ACT_SITE 1063 1063 BY SIMILARITY.
 FT MOD_RES 902 902 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 Query Match 28.4%; Score 1932; DB 1; Length 1426;
 Best Local Similarity 32.6%; Pred. No. 1e-95;
 Matches 444; Conservative 183; Mismatches 439; Indels 338; Gaps 41;
 QY 24 QVCTGDMKRLPASPETHLMDRLVGGCVQVGNLETLVPTNASLSTLODIOEVQ 82
 DB 100 KICIGTKRSLSPVSKNHHYRLRLRYNTCYVDNLTWLPENMDLSFLDIRVETG 159
 QY 83 YIKANSKFGITELRLRLRIVRGTOLEF-----EDNYALAVLNGDPLNNTVTGASPGCL 137
 DB 160 YILISHVDVKKVYVFKLOIRRLTFLSVSEBKALV-----TYSKM 203
 QY 138 RELQRLSTLEIKGVLIQRPDLQCYQDTIIMKDIHKNQALTLDTNSRACHPCSP 197
 DB 204 YLLEIPDLRDVLNGQVGFNNYNLCHMRTIQWSEIVSNGTAYVYDFTADRECPKCHE 263
 QY 198 MCKGSRGWGESESDQSLRTVYACAGCA--RCKGPLPTDCCHEGCAAGCTGPKXSDCLAC 255
 DB 264 SCTHG-CWEGEPKCKCKSKLTCSPQCAAGRCYGPXKPECCHLFGAGGCTGPTQKDCIAC 322
 QY 256 LPHNSGICELCPALVTYNTDFESMNPREGRYTFGASCTYACPYNYLSTDVSGCTLVLC 315
 DB 323 KNFFDAVSKCECPKPKRYNPTTYVLETPBEKVAYGATVCECP-GHLRDNGACVASC 381
 QY 316 PLHNGVTLLEDGTQCEKSKCAVCGLGHEHLREVRATYSANIQFPAACKYIFGLA 375
 DB 382 PDKKMDKGE-----CVPNGCPPTCGVTVLH-----AGTDSFRNCTYIDGIR 428
 QY 376 FLPESPDG--DPASNTA-----PLQPEQLQVFETLEITGLYLSAMPDLPPLSVFQN 427
 DB 429 ILDQTSGFQDYAVNTWMPRYTLPDPERREVFVSKITGLVINEGTHPQFRLSYERN 488
 QY 428 LQYIRGRITLHNGAY-SLTQIGIGIWLGLSLRLSGSLALIHNTHLCPHYTYVMDOLF 486
 DB 489 LETIRGRQLMESMFALALVKSLSVLSMRNLKQISSSVVIOHNRDLCYVSNIRWPAIQ 548
 QY 487 RNPQALLHTANRPEDECVGEGLAQHCARQHGCPPTGCVNCSQPLRGOECVBEGRV 546
 DB 549 KEPEQKVMNENLRADLCXKNGTITSDQCNEDGCGAGATDQCLTKNFRNGTCLADCGY 608
 QY 547 LOGLPREYVNAHHCIPCHPECOFNGSVTCFGEPAADCCVACAHYKDPFCVACRP----- 601


```

Db      609 ISNAK--FDNRKTCIKHPECR-----TCNAGAGADHCQECVHVRGQHCVSBCPNKYN 660
Qy      602 -----SGVK-----PDLSTYPIW 614
Db      661 DRGVRECHATCDGCTGPKDTIGTIGACTTCNLAIINNDAIVAKGLKDKDCPD-GI--FW 717
Qy      615 KF--PDEBGACQP-----CPI-----NCTH-----632
Db      718 EYVHPOEGSLKPLAGRAVCRKCHPLCELTNYGHEQVCSKCTHYRRRECECECPADH 777
Qy      633 -----SC-----VDLDKXG-----CPAD 646
Db      778 YTDGEQCECFORHECECGCTGPGADCKSCRNFKLPDANETGPPYVSTFNCTSKCPLEW 837
Qy      647 R-----ASPLTS-----YSAAVGLLVVVLGVVFGILIKRQ 679
Db      838 RHNVYQYTAIGPYCAASPRSSKITANLDVNMIFITGAVALVPTICLCV--TYICRQK 895
Qy      680 QKIRKVT--MRLLQETELVEPLTPSGAMNQAMLIKETELRKVKVIGSGAFGVYK 737
Db      896 QKAKKEIVKMTMALSGEDSEPLRPSITGNLCKLRIVKDAELRKGVLMGAFGRVYK 955
Qy      738 IWPDGENVKIPVAIKVLRNTSPKANKELDEAYVMAVGSPYVRLGICLTSTVOLV 797
Db      956 VMVEGENVKIPVAIKELKSTGASESEEFLEAYIMASEHEHNLKLAVCMSSQMMLI 1015
Qy      798 TQLMPYGCLLDVHRENGRLGSDLLNMCQIAKMSYLEDVRLVFRDIAANVLKSPN 857
Db      1016 TQLMPLGCLLDVYANNNDKIGSKALNWSQIAKMSYEEKLVHARDIAANVLYQTPS 1075
Qy      858 HVKTTDFGLARLLDIDETEHADGKVPKIMMALESILRRFTHSDVMSYGVWELMT 917
Db      1076 LVKITDFGLAKLLSSDSNEKYKAGKMPIMMLIECTRNRFISKDVAFGVITWELT 1135
Qy      918 FGAKPYDGPAREIPDLLEKGERLPPPICTIDVIMVYKCMWIDSECPREELVSEPS 977
Db      1136 FGQRPHEINIPAKDIPDLIEVGLKLEQFEICSDIYCTLLSCWMLDAMAPTKQTLTVA 1195
Qy      978 RMARDDPQRFVIGNEDLG--PASPLDSTFRSLLED--DMQDLVAEYLVPQGFRC 1032
Db      1196 EFARDPGRYALIEGDKTRLP-----YTSODEKDLIRKLAFTTDSSEALAKDDYLQ 1248
Qy      1033 PDPAFGAGMVRHRRSSSTRSGGDLTLGLEPSEBAP-----RSLPAPSEAGSDVF 1086
Db      1249 PKAAPGPS-----HRDCT-----DEMEXLNRKCKDPSNKKSSTGDDER 1287
Qy      1087 DG---DLGMAKAGLQSLPHDPSPLQRISEDPVTLBSETGTGYAPLTCSQPBYVNP 1143
Db      1288 DSSAREVGVGNLR-----IDLPEDEDDYLMF--TCQPGPNNNNMM 1325
Qy      1144 DVRRPPSPREPLPAPAPAGATLERAKTILSPGKGVKCVAFAGAVENPEYL-----TP 1199
Db      1326 N-----NPNQNMMAVAVGAYM-----DLGVSVSDNPEYILNAQTL 1364
Qy      1200 QGGAAPQPH-----PPAPFSP-AFDNLVYWD 1224
Db      1365 GVGESPIPTGTTIGIPYMGPGTVEVKVPMRGSEPTSSDHEVYND 1408

```

RESULT 12

```

ERBB ALV
ID ERBB ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN v-ERBB
OS Avian leukosis virus.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
NCBI_Taxid=11864;
RN [1]

```

```

RP      1 SEQUENCE FROM N.A.
RA      2 MEDLINE=85228222; PubMed=2988784;
RX      3 Nilsen T.W., Maroney P.A., Goodwin R.G., Rotman F.M.,
RA      4 Crittenden L.B., Raines M.A., King H.-J.,
RT      5 "c-erbB activation in ALV-induced erythroid blastosis: novel RNA
RT      6 amino-truncated EGF receptor.";
RL      7 Cell 41:719-726(1985).
CC      8 -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      9 tyrosine phosphate.
CC     10 -1- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC     11 IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC     12 AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC     13 THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC     14 C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC     15 -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC     16 PROTEIN.
CC     17 -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC     18 -----
CC     19 This SWISS-PROT entry is copyright. It is produced through a collaboration
CC     20 between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC     21 the European Bioinformatics Institute. There are no restrictions on its
CC     22 use by non-profit institutions as long as its content is in no way
CC     23 modified and this statement is not removed. Usage by and for commercial
CC     24 entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC     25 or send an email to license@isb-sib.ch).
CC     26 -----
CC     27 EMBL; M10066; AAA48763.1; ALT_INVIT.
CC     28 PIR; A00643; TVCHLV.
CC     29 PIR; B00643; TVPVLV.
CC     30 HSP; P11362; 1FGK.
CC     31 InterPro; IPR000719; Euk_pkinase.
CC     32 InterPro; IPR001245; Tyr_pkinase.
CC     33 Pfam; PF00069; pkinase; 1.
CC     34 PRINTS; PR00109; TYRKINASE.
CC     35 ProDom; PD000001; Euk_pkinase; 1.
CC     36 SMART; SM00219; TyrcKc; 1.
CC     37 PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC     38 PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC     39 PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC     40 TRANSFERASE; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC     41 Glycoprotein; Phosphorylation.
CC     42 DOMAIN 132 399
CC     43 NP_BIND 138 146
CC     44 BINDING 165 165
CC     45 ACT_SITE 257 257
CC     46 BY SIMILARITY.
CC     47 SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;
CC     48 -----
CC     49 Query Match 25.7%; Score 1749.5; DB 1; Length 634;
CC     50 Best Local Similarity 52.3%; Pred. No. 2.2e-86;
CC     51 Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;
CC     52 -----
Qy      587 CAHYKDPFCVACPSGVKPDLSYPMKFPDEBACQPCPINTCHSCVLDLDDKCPAQ 646
Db      3 CAHFDGPHCVACPAVGLGENDTL-VWKYADANVVCQCHNCRGCKGGLGECF--- 58
Qy      647 BASPLTSISAVV-GILVVY-GVVFGLIKRQKIRKTYMRLQGTETELVEPLTPSGA 705
Db      59 NSKSTPSIAGVAVGGCLLVAVGLGIGYLRRR-HIVKRTLRLQLQREELVEPLTPSG 117
Qy      706 MPNQAMRIKETELRKVKVYLSGAFGVYKGIWIPDENYKIPIAIVLRNTSPKANK 765
Db      118 APQAHRLIKETEFKXVYVLSGAFGVYKGIWIPBEKIKIPAIKELNKAISPRANK 177
Qy      766 ELIDENYVMAVGSPYVRLGICLTSTVOLVLTQMPYGCILDVHRENGRLGSDLLNW 825
Db      178 ELIDENYVMAVDNPHVCRLLGICLTSTVOLVLTQMPYGCILDVIREHKDNGISQYLNN 237
Qy      826 CQIAKMSYLEDVLRDIAARVAVLKSPPHYKTTDFGLARLLDIDETEHADGKVP 885
Db      238 CQIAKMSYLEDVLRDIAARVAVLKSPPHYKTTDFGLARLLDIDETEHADGKVP 297
Qy      886 IKMALESILRRFTHSDVMSYGVWELMTFGAKPYDGPAREIPDLLEKGERLPP 945

```

```

DB 298 IKMALESLHRRITTHQSDVWSYGVYTWELMTFSKRYDGPASEISSVLEKGERLPQRP 357
QY 946 ICTIDVYMIWVKCMWIDSECRPRERELVSFSESMARDPQRFVYIQ-NEDUGPASPUDSTF 1004
DB 358 ICTIDVYMIWVKCMWIDADSRPKRELIAFSKWARPPRYLVYQGERMHLSPDTSKF 417
QY 1005 YRSLLDDMDGLDYDAEYLYVPOGFCPCPPAPAGAGMWHRRSSSTRSGGDLTLGLE 1064
DB 418 YRTLMEEDMEDIVDAEYLYVPHQGF-NSPST 449
QY 1065 PSEEPAPSP-APSEGAGSDVFDGDLGMAAGLQSLPTHPSPLQRYSEDPYV 1119
DB 450 -----SRPLLSLSATSNNSAATNCID-----RNGGCHVREDSPYQRRSSPTGN 495
QY 1120 LPSET--DGYVAPLTGSPQPEYVNPQVPQPSPREGRLPAARPAATLTERAKTISPCK 1177
DB 496 FLEESIDGFL-----PAPRYVNO-LMPKKPS-----TAMQ 526
QY 1178 NGVVKDF-----AFGAVENPEYLTPOGGAAPQHPPPAFSPAFDNLVY 1222
DB 527 NQYNNISLTALSKLPMDSRYONSHSTAVDNPYL-----NTNGSLAKTVFESSPY 578
QY 1223 WDQ-----DPEP-----KAPSTFKGPTAENPYLGLDVP 1254
DB 579 WIQSGNQINLNDPQODFLPNETKPMGLKVPAAENPEYLRVAAP 625

RESULT 13
ERBB_AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (BC 2.7.1.112).
CN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN 11
RC STRAIN=H;
RP SEQUENCE FROM N.A.
RX MEDLINE=84026539; Pubmed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbb gene of avian erythroblastosis virus is a member of the src
RT gene family.";
RL Cell 35:71-78(1983).
RN 12
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; Pubmed=6328659;
RA Deblie B., Henry C., Benaisa M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erba gene of avian erythroblastosis virus reveals a
RT new type of oncogene.";
RL Science 224:1456-1459(1984).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -1- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1; -.
DR PIR; A00644; TYUH.
DR HSSP; P11362; IFK.
DR InterPro; IPR000719; Euk PKinase.
DR InterPro; IPR001245; Tyr PKinase.
DR Pfam; PF00069; Pkinase; 1.
DR ProDom; PD000001; Euk PKinase; 1.
DR SMART; SM00219; TykC; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE; PS00111; PROTEIN KINASE DMR; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76EBDD06745D609 CRC64;

Query Match 25.0%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 6,4e-84;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

QY 587 CAHYKDPFCVARGSGVGRDLSMPYMKRPFDEGACOPFCINCTHSVCLDDKGGPAG 646
DB 3 CAHFTDGHCVKACPAVLGENDTL-VKRYADANAVQCLHPCTGCKGPGLEGCP--- 58
QY 647 RASPLTSIVSAV-GILLVVLGVVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGA 705
DB 59 NGSKTPIAAVGVGGLLVVVGIGLYLRR-HYRKXLTLLRLQERLVEPLTPSGE 117
QY 706 MPNQAKELKETLARKVYLGSAFGTVKGIWIPDGENVKIPLVAKVURENTSPYANK 765
DB 118 APQAHRLKETETFFKVKYVLSGSAFSTIKGIMPEGEVKYKIPVAKELREATSPYANK 177
QY 766 EILDEAYMAGVSPYVSRILGICLTSTVQVLTQVLMYGCGLDHVENRGLSGDILNM 825
DB 178 EILDEAYMASVDNPHYCRLLIGLICITVLTQVLMYGCGLDYIEHKNIQSQVLLNM 237
QY 826 CMQIAKGSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLDIDETEVHADGKVP 885
DB 238 CVOIAKGMNLYEERLVRDLAARNVLKTPQVYKTFGLAKILGADKEXYHAEGKVP 297
QY 886 IKMALESLHRRITTHQSDVWSYGVYTWELMTGAKRYDGPASEISSVLEKGERLPQRP 945
DB 298 IKMALESLHRRITTHQSDVWSYGVYTWELMTGSKRYDGPASEISSVLEKGERLPQRP 357
QY 946 ICTIDVYMIWVKCMWIDSECRPRERELVSFSESMARDPQRFVYIQ-NEDUGPASPUDSTF 1004
DB 358 ICTIDVYMIWVKCMWIDADSRPKRELIAFSKWARPPRYLVYQGERMHLSPDTSKF 417
QY 1005 YRSLLDDMDGLDYDAEYLYVPOGFCPCPPAPAGAGMWHRRSSSTRSGGDLTLGLE 1064
DB 418 YRTLMEEDMEDIVDAEYLYVPHQGF-NSPST 449
QY 1065 PSEEPAPSP-APSEGAGSDVFDGDLGMAAGLQSLPTHPSPLQRYSEDPYV 1119
DB 450 -----SRPLLSLSATSNNSAATNCID-----RNGGCHVREDSPYQRRSSPTGN 495
QY 1120 LPSET--DGYVAPLTGSPQPEYVNPQVPQPSPREGRLPAARPAATLTERAKTISPCK 1177
DB 496 FLEESIDGFL-----PAPRYVNO-LMPKKPS-----TAMQ 524
QY 1178 NGVVKDF-----AFGAVENPEYLTPOGGAAPQHPPPAFSPAFDNLVY 1218
DB 527 NQYNNISLTALSKLPMDSRYONSHSTAVDNPYL-----NTNGSLAKTVFESSPY 574
QY 1219 NLVYWDQDPEPARGAPSTFKGPTAENPEY 1248

```

Db 575 S8PYWISQGNHQ-----INLNDPXY 594

RESULT 14
ERBB_AVIEU
ID ERBB_AVIEU STANDARD; PRT; 540 AA.

AC P11273;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN v-ERBB.
OS Avian erythroblastosis virus (strain ts167).
OC Viruses; Retrocid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=103898;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=87064458; PubMed=2878364;
RA Choi C.R., Trainor C., Graf T., Beug H., Engel J.D.;
RT "A single amino acid substitution in v-erbB confers a thermolabile
RT phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT cells."
RL Mol. Cell. Biol. 6:1751-1759(1986).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC -----
CC EMBL: M13179; AAA2401.1; .
CC PIR: A25231; TVFEB.
CC HSP: P11362; IFGK.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR001245; Ty_pkinase.
CC Pfam: PF00069; pkinase; 1.
CC DR ProDom: PD000001; Euk_pkinase; 1.
CC DR SMART: SM00219; TyKc; 1.
CC DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
CC DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
CC KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC Glycoprotein; Phosphorylation.
CC FT DOMAIN 132 399
CC NP_BIND 138 146 ATP (BY SIMILARITY).
CC FT BINDING 165 165 ATP (BY SIMILARITY).
CC FT ACT_SITE 257 257 BY SIMILARITY.
CC FT VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).
CC SEQUENCE 540 AA; 60412 MW; 5B53297AA068B65D CnC64;

Query Match 23.9%; Score 1630; DB 1; Length 540;
Best Local Similarity 55.1%; Pred. No. 4.4e-80;
Matches 341; Conservative 69; Mismatches 119; Indels 90; Gaps 14;

QY 587 CAHYKDPFCVACRSGGKVPDLSTYPIKPFDEGACQPCPCINCHGCVLDIDKCPAQ 646
Db 3 CAHFDGPHCVACAGVAGLNDLTVKADANAVCOLCHPNTCKCPGEGCP--- 58
QY 647 RASPLTSIVSAVV-GILVVVVGAVFGILIRROOKTRKYMRLLOETELVEPLPSGA 705
Db 59 NGSKTPSIAAGVVGGLLVVVGIGLYLRR-HYVKRLRLDLERLVEPLTPSGE 117
QY 766 MPNOQMRILKETELRKRYVCSGAFGVYGVIVPDGENVKIPVAIKVIRENTSPYANK 765
Db 118 AFNOHLRLILKETELRKRYVCSGAFGVYGVIVPDGENVKIPVAIKVIRENTSPYANK 177
QY 766 EILDEAYVWAGVGSYVSRLLGLCTSTVQVLVLTQMPGCLLDVVRNRRLQSODLNN 825

Db 178 EILDEAYVWASVDNPHVCRLLGLCTSTVQVLITQMLPYGCLLDYIREKHONIGSQYLLNW 237
QY 826 CMQIAKMSVLEDEVLVRLAARVVLVKSNNHYKITDFGLARLLDIDETVHADGGVVP 885
Db 238 CVQIAKMSVLEDEERHVRDLAARVVLVKTQHYKITDFGLAKQGADEKEYHABGGVVP 297
QY 886 IKMMALESILRRRPTHQSDVMSYGVTVMLTFGAKPYDGIPIAREIPDLLEKGRLLQPP 945
Db 298 IKMMALESILHRIYTHQSDVMSYGVTVMLTFPGSKPYDGIPIAREISVLEKGRLLQPP 357
QY 946 ICTIDVYIMVWKMTDSECPREFVELSESRKARDPQRFVVIQ-NEDLGPAPLSDTF 1004
Db 358 ICTIDVYIMVWKMSDADSPREFELIAEFSKARPPRYLVIQGERMLPSPDTSKF 417
QY 1005 YSLDEDDMDGLVDAEYVLPQOQFPCPDAPAGVGHRRHSSSTRSGGDTLIGLE 1064
Db 418 YRLTMEEMEDIVDAEYLVPHQCF-----NSPST----- 449
QY 1065 PSEEDAPRSPF-----APSEAGSDVFDGDLGMAAKGLQSLPHDPSPLQYSEDPTVP 1115
Db 450 -----SRTPLLSLSATSNNSATVCIDRNG-----H----- 476
QY 1120 LPSRDGVVAPITCSPOPEYVNOPDVRQPSREGPLAPAPAGAT-LEKAKTSPGKN 1178
Db 477 -PVREDGFL-----PAEYVNO-LNFKKPSYAMVQNYVISTALISKLPIDSRYQN 527
QY 1179 GVYKDVFAFGAVENPEYL 1197
Db 528 -----SHSTAVDPEYL 539

RESULT 15
EGFR_CHICK
ID EGFR_CHICK STANDARD; PRT; 703 AA.

AC P13387;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER
DE (Fragment)).
GN EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=88261272; PubMed=3260329;
RA Lax I., Johnson A., Hawk R., Sap J., Bellor F., Winkler M.,
RA Ulrich A., Vennstrom B., Schlessinger J., Givol D.;
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha."
RL Mol. Cell. Biol. 8:1970-1978(1988).
CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF.
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

or send an email to license@sib.ch).

```
CC -----
DR EMBL; M20386; AAA48760.1;
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP_PARTIAL.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR_PARTIAL.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM_PARTIAL.
KM Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 30
FT CHAIN 31 703
FT DOMAIN 31 654
FT TRANSMEM 655 667
FT DOMAIN 668 >703
FT DISULFID 197 206
FT DISULFID 201 214
FT DISULFID 222 230
FT DISULFID 226 238
FT DISULFID 239 247
FT DISULFID 243 255
FT DISULFID 258 267
FT DISULFID 271 298
FT DISULFID 302 314
FT DISULFID 316 333
FT DISULFID 316 340
FT DISULFID 316 352
FT DISULFID 517 530
FT DISULFID 533 542
FT DISULFID 546 562
FT DISULFID 565 581
FT DISULFID 589 592
FT DISULFID 592 601
FT DISULFID 605 627
FT DISULFID 630 638
FT DISULFID 634 646
FT CARBOHYD 134 134
FT CARBOHYD 190 190
FT CARBOHYD 200 200
FT CARBOHYD 200 200
FT CARBOHYD 359 359
FT CARBOHYD 368 368
FT CARBOHYD 420 420
FT CARBOHYD 573 573
FT CARBOHYD 578 578
FT CARBOHYD 613 613
FT CARBOHYD 633 633
FT CARBOHYD 648 648
FT NON_TER 703
SQ SEQUENCE 703 AA; 77427 MW; AFE2DE11B735A690 CRC64;

Query Match 23.3%; Score 1586; DB 1; Length 703;
Best Local Similarity 43.6%; Pred. No. 1,3e-77;
Matches 306; Conservative 115; Mismatches 256; Indels 28; Gaps 12;
```

```
QY 8 RWGLLALLPPGA-----STQVCTGTDKRLDPASEPTHLDMLRHLVYQGGQVYVGNLE 61
DB 13 RGAAYLVLLLGVALCSAVEEKVKVCOGNNKLTQLGHVEDHFTSLQRYNNCEVALSNLE 72
QY 62 LTYLFTNASTLSFDIPIQEVQYIKANSKFITGTELQRLRIVRGTLFEDNYALAVLNDG 121
DB 73 IYVEHNRDLTFELKTIQVAGYVLLALNVDYIPLENLQIRGNVLYDNSFALVLSNYH 132
QY 122 PLNNTPVTGASPGGLREIQRLSLTEILKGVLLIQRNPLCYQDTILMKDI FHKNNQLAL 181
DB 133 -MVKIQ-----GIRELPMKRLSEILNGVKISNNPKLCNMDIVLWNDIIDTSRK-PL 182
QY 182 TLID-TNRSRACHPSGMCKSGSRGWGESEDDQSLTRTVCAAGCA-RCKGSLPTDCCHQ 239
```

```
DB 183 TVLDFASNLSSCPKHPCTDEHDHGMAGSEQNCQTLTVKICAQCCSGRCRGKVPSPDCCHQ 242
QY 240 CAAGCTGPKASDCLACHFNHSGICEIHCAPALVYNTDTESNNEPGRITTFGASCYTAC 239
DB 243 CAAGCTGPRESDCLACRFDDACCKTCPCPLVYNTTYQMDVNPBGKYSFGATCVREC 302
QY 300 PNYLSTVGSCTVLCPLHNOEVTAEADOTGRCESGSPCARVCYGLMEHLREYRAVTS 359
DB 303 PNYVTVTHGSCVNSCNTDYEV-ENGVKRCCKGCDGLCSKVCNGIGTIGELKGLSINAT 361
QY 360 NIOEPAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVETLEETIGYLYISAWPDSL 419
DB 362 NIDSKNCKTKINGDVSILPVAFLDPAFKTLPDPKXLDVFRVKEISGFLLIQAMPDNA 421
QY 420 PLSLFGQLQVIRGILHNGAVSLTLQGLGSMWGLSLRSLGSLAIHHNTHLCVHT 479
DB 422 TDLVAFENLEIRGTXKHGOYSLAVNKLQSLGLSLKEISDGDIAIMKNKMLCYADT 481
QY 480 VPMDFLFPNPHQALHTANREDEECVGEGLACHQLCARGHCMGPGPTQCVNCSQFLRGQE 539
DB 482 MMWRSLFATQSKTKIIGNRNKNOCTADRHVCDPLGSDVGCMGPGPHCTSCRFSSRQKE 541
QY 540 CYEGRVYQGLPREYVNAHCLPCHPBCQPNQ---SYTCRFPADOCVACATKDPFC 596
DB 542 CVKQCNILQGEPRERERDSKCLPCHSECLVONSTAYTTCSGPGPDHCKMCAHFDGPHC 601
QY 597 VAPCSGVKPLDSYMPIMKFPDEEGACQPCINCTHSCVDLDBKGCFAEORASPLTSIVS 656
DB 602 VVACAGVLTGENDTL-VKRYADAANVQGLCHPNTTRCGKRGLEGCP---NGSITPSIAA 657
QY 657 AVV-GILLVVLGVVFGILLRQOKIRKRYMRRLQETELVEPLTP 702
DB 658 GVVGGLCLVWVGILGILYLRH-HIVKXTRLRLQERLVEPLTP 703
```

Search completed: July 22, 2003, 08:44:03
Job time : 20.2504 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.3575 Seconds

(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-59-73-12

Perfect score: 6809

Sequence: 1 MEALALCRWGLLALLPPGA.....TFKGFPAENPEYIGLDPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP archaea: *
2: SP bacteria: *
3: SP fungi: *
4: SP human: *
5: SP invertebrate: *
6: SP mammal: *
7: SP mhc: *
8: SP organelle: *
9: SP phage: *
10: SP plant: *
11: SP rodent: *
12: SP virus: *
13: SP vertebrate: *
14: SP unclassified: *
15: SP viirus: *
16: SP bacteriap: *
17: SP archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6223	91.4	1259	6	018735
2	3120	45.8	1209	11	090X70
3	3091	45.4	1210	11	09EP98
4	2729.5	40.1	1137	13	09W6F6
5	2703	39.7	1165	13	09YH40
6	2275	33.4	1328	13	P79754
7	2011.5	29.5	1433	5	09B1H9
8	1797.5	26.4	419	4	09UK79
9	1739	25.5	367	11	08R2X1
10	1720	25.3	729	15	086712
11	1718	25.2	567	15	086714
12	1697.5	24.9	412	4	08WYV0
13	1653.5	24.3	962	15	064895
14	1645	24.2	545	15	085468
15	1482.5	21.8	655	11	09WV75
16	1466.5	21.5	643	11	09ERV6

17	1274	18.7	1193	5	09Y1X8	09Y1X8 ephydratia f
18	1173.5	17.2	1368	5	023821	023821 ctenorhabdi
19	1146	16.8	1217	5	026565	026565 schistosoma
20	1117	16.4	527	13	090836	090836 gallus gall
21	982.5	14.4	478	11	09SE50	09SE50 rattus norv
22	933.5	13.7	599	13	09PSH2	09PSH2 gallus gall
23	906	13.3	165	4	014256	014256 homo sapien
24	887	13.0	176	11	0233V5	0233V5 rattus norv
25	806.5	11.8	345	13	P11776	P11776 xiphophorus
26	778	11.4	435	5	0862M1	0862M1 xiphophorus
27	754.5	11.1	311	13	09P162	09P162 xiphophorus
28	753.5	11.1	1362	13	09PV24	09PV24 xenopus lae
29	732	10.8	1671	5	09NJV5	09NJV5 biemphalari
30	723	10.6	149	6	09BG66	09BG66 oxytelagus
31	702	10.3	1418	13	093457	093457 scophthalmu
32	701.5	10.3	1368	13	08W65	08W65 paralicthy
33	692	10.2	331	4	09BUD7	09BUD7 homo sapien
34	678.5	10.0	1369	13	08W66	08W66 paralicthy
35	678	10.0	1358	13	073798	073798 xenopus lae
36	658	9.7	1412	13	08W64	08W64 paralicthy
37	655.5	9.6	1472	5	09U5A6	09U5A6 bombyx mori
38	646.5	9.5	1418	13	08W63	08W63 paralicthy
39	645	9.5	1245	13	09YGH8	09YGH8 scophthalmu
40	642	9.4	1571	11	09QVW4	09QVW4 rattus sp.
41	617.5	9.1	2144	5	09VD94	09VD94 drosophila
42	600	8.8	935	4	096135	096135 homo sapien
43	599	8.8	987	11	091YMO	091YMO mus musculu
44	596	8.8	987	11	099MR2	099MR2 mus musculu
45	587.5	8.6	1036	4	007912	007912 homo sapien

ALIGNMENTS

RESULT 1

018735

ID 018735

AC 018735

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Erbs-2

OS Canis familiaris (Dog)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;

OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

OX NCBI_TaxID=9615;

RN [1]

RP SEQUENCE FROM N.A.

RA Yokota H.

RT "CDNA cloning of erbs-2 from canine mammary gland."

RL Submitted (Oct-1997) to the EMBL/Genbank/DBJ databases.

DR EMBL; AB008451; BAA2127.1; -

DR HSSP; P11362; 1FGR.

DR InterPro; IPR002048; EF-hand.

DR InterPro; IPR000494; EGFR_L domain.

DR InterPro; IPR000719; Euk_Pkinase.

DR InterPro; IPR002174; Furin-like.

DR InterPro; IPR001245; Tyr_Pkinase.

DR InterPro; IPR004019; YLP_motif.

DR Pfam; PF00757; Furin-like; 1.

DR Pfam; PF00069; Pkinase; 1.

DR Pfam; PF01030; Recep_L_domain; 2.

DR Pfam; PF02575; YLP_2.

DR ProDom; PD000001; Euk_Pkinase; 1.

DR SMART; SM00261; FU; 3.

DR SMART; SM00219; TYRC; 1.

DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.

DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

DR ATP-binding; Transferase; Tyrosine-protein kinase.

SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

Query Match 91.4%; Score 6223; DB 6; Length 1259;
 Best Local Similarity 91.2%; Pred. No. 0;
 Matches 1149; Conservative 39; Mismatches 66; Indels 6; Gaps 2;

QY 1 MELAALCRWGLLALLPGAASTQVCTGDMKLRPASPETHLDMLRHL YGQCQVVGNTL 60
 DB 1 MELAAACRWGLLALLPGSAGTQVCTGDMKLRPASPETHLDMLRHL YGQCQVVGNTL 60
 QY 61 ELTYLPTNASLSTLQIQEVQOYIKANSKFIGITELQRIIRGTQLEEDNYALAVDNG 120
 DB 61 ELTYLPTNASLSTLQIQEVQOYIKANSKFIGITELQRIIRGTQLEEDNYALAVDNG 120
 QY 121 DPLNNTPTVGTASPGSLRELOLRSLTEILKGGVLIQRPOLCYQDTIIMKDI FHKMQLA 180
 DB 121 DPLNNTPTVGTASPGSLRELOLRSLTEILKGGVLIQRPOLCYQDTIIMKDI FHKMQLA 180
 QY 121 DPLNNTPTVGTASPGSLRELOLRSLTEILKGGVLIQRPOLCYQDTIIMKDI FHKMQLA 180
 DB 121 DPLNNTPTVGTASPGSLRELOLRSLTEILKGGVLIQRPOLCYQDTIIMKDI FHKMQLA 180
 QY 181 LTIIDTNRSRACHPGSPMCKSGSRCKWSESEDCQSLTRTYCAGGACRCKGRLPTDCHEQC 240
 DB 181 LTIIDTNRSRACHPGSPMCKSGSRCKWSESEDCQSLTRTYCAGGACRCKGRLPTDCHEQC 240
 QY 241 AAGCTGPKHSDCLACHFNHSGICEIHCALVTYNTDTFESMPNEGRYTFGASCVTACP 300
 DB 241 AAGCTGPKHSDCLACHFNHSGICEIHCALVTYNTDTFESMPNEGRYTFGASCVTACP 300
 QY 301 YNYLSTDVGSCTLVCPFLANOVTAEPTQRCCKSPCARVCTGYLGMHLLREVTASAN 360
 DB 301 YNYLSTDVGSCTLVCPFLANOVTAEPTQRCCKSPCARVCTGYLGMHLLREVTASAN 360
 QY 361 IOEFAGCKKIFGSLATLPESFDGDPASNTAPLOPELOVFEETLEETLYLYISAMPDLP 420
 DB 361 IOEFAGCKKIFGSLATLPESFDGDPASNTAPLOPELOVFEETLEETLYLYISAMPDLP 420
 QY 421 DLSVFQNLQYIRGRILHNGAYSLTLOGIGISWLGRLSLEIGSGALIHNTLCFVATV 480
 DB 421 NLSVFQNLQYIRGRILHNGAYSLTLOGIGISWLGRLSLEIGSGALIHNTLCFVATV 480
 QY 481 PMDQLEFRNPQALHTANRPEDECEGEGIACHQOLCARGCMWRGPTQCVNCSQFLRGEC 540
 DB 481 PMDQLEFRNPQALHTANRPEDECEGEGIACHQOLCARGCMWRGPTQCVNCSQFLRGEC 540
 QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCAVAXKDPPEVCARC 600
 DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCAVAXKDPPEVCARC 600
 QY 601 PSGVKPDLSTMPWKRPDEEGAQPCPINCTHSQVDLDKXGPAQORAPLTSTYSANVG 660
 DB 601 PSGVKPDLSTMPWKRPDEEGAQPCPINCTHSQVDLDKXGPAQORAPLTSTYSANVG 660
 QY 661 ILLVVLGVVFGILIRROQKIRKYMRLLOETELVEPLTPSGAMPQAOQRILKETEL 720
 DB 661 ILLVVLGVVFGILIRROQKIRKYMRLLOETELVEPLTPSGAMPQAOQRILKETEL 720
 QY 720 RYKXVVGSGAFGVYVGIWIPDENNKIIVAKVIRENTSPANKKEIIDEAUVAGVSP 779
 DB 720 RYKXVVGSGAFGVYVGIWIPDENNKIIVAKVIRENTSPANKKEIIDEAUVAGVSP 779
 QY 781 YVSRLLGICLTSVQVLTQMPYGCCLLDHVRNRRGLSQDILNMCQIAKMSYLEQVR 840
 DB 781 YVSRLLGICLTSVQVLTQMPYGCCLLDHVRNRRGLSQDILNMCQIAKMSYLEQVR 840
 QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETRYHADGKVPDKMMALESILRRFT 900
 DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETRYHADGKVPDKMMALESILRRFT 900
 QY 901 HOSDWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGRRLPOPICTIDYVMIVKMM 960
 DB 901 HOSDWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGRRLPOPICTIDYVMIVKMM 960
 QY 961 IDSECRPRFRELVSFSRWARDPQRFVVI QNEDLGPASLDSTFYRSLLDDMDGLVDA 1020
 DB 961 IDSECRPRFRELVSFSRWARDPQRFVVI QNEDLGPASLDSTFYRSLLDDMDGLVDA 1020
 QY 1021 EBYLVPOQGFPCDPAPAGAGVYHHHRSSSTRSGGDLTLGLEPSEEBAPRSLAPSEG 1080

DB 1020 EBYLVPOQGFPCDPAPAGAGVYHHHRSSSTRSGGDLTLGLEPSEEBAPRSLAPSEG 1075
 QY 1081 AGSDVFDGDLGMAKAGCQSLPTHDPSPLOQYSEDPYPLPSETFDGYAPLTCSPQPEYV 1140
 DB 1080 AGSDVFDGDLGMAKAGCQSLPTHDPSPLOQYSEDPYPLPSETFDGYAPLTCSPQPEYV 1135
 QY 1141 NOPDYAPQPPSPREGPLPAPAPAGATLRR-----AKTSPGKGVYKDVFAFGAVENPE 1195
 DB 1140 NOPDYAPQPPSPREGPLPAPAPAGATLRR-----AKTSPGKGVYKDVFAFGAVENPE 1195
 QY 1196 YLTPQGAAPQHPAPPAPSPAFDNLVYWDQDPERKADPSTFKGTPTAENDEYGLDVPV 1255
 DB 1200 YLAPRGRAAPQHPAPPAPSPAFDNLVYWDQDPERKADPSTFKGTPTAENDEYGLDVPV 1255

RESULT 2
 Q9QX70 PRELIMINARY; PRI: 1209 AA.
 AC Q9QX70;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Epidermal growth factor receptor.
 GN EGFR.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RX MEDLINE=90258688; PubMed=2342466;
 RA Patch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,
 RT "A truncated, secreted form of the epidermal growth factor receptor is
 encoded by an alternatively spliced transcript in normal rat tissue."
 RM Mol. Cell. Biol. 10:2973-2982(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RA Patch L.A.;
 RL Submitted (NOV-1991) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RA Guttridge K., Dawson T.L., Earg H.S.;
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: M37394; AAF14008.1; -.
 DR HSSP: P11362; 1RGX.
 DR InterPro: IPR000494; EGFR_L domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR01245; Tyr_pkinase.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00669; pkinase; 1.
 DR Pfam: PF01030; Recept_L domain; 2.
 DR PRINTS: PRO0109; TYRKINASE.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; FU; 3.
 DR SMART: SM00219; TYRK; 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE: PS50011; PROTEIN KINASE DOM; 1.
 DR PROSITE: PS00109; PROTEIN KINASE_TYR; 1.
 KM APP-binding; Receptor; Transferrase; Tyrosine-protein kinase.
 SQ SEQUENCE 1209 AA; 134891 MW; 56FE8786C1B773 CRC64;

Query Match 45.8%; Score 3120; DB 11; Length 1209;
 Best Local Similarity 49.6%; Pred. No. 4-4e-228;
 Matches 634; Conservative 167; Mismatches 362; Indels 114; Gaps 26;

QY 3 LALACRWGLLALLPGAASTQVCTGDMKLRPASPETHLDMLRHL YGQCQVVGNTL 61

Db 15 LALCAAG-----GALEEKVCQGTSTNRILTQIGTEFEDHFLSLQEMFNNECVLLGNLE 66
 QY 62 LTYLPTNASTLFLQDIOEVOQYIKANSKFIGITTELORLIRVGTOLPEDNYALAVLNGD 121
 Db 67 ITYVQRNDLSFLKTIQEVAGYVLLALMTVERIPIENQIIRGNALVENTYALAVLEN-- 124
 QY 122 PLNNTPTVTGASPGGLRELOLSLTELKGVLLIQRNFQOLCYQDTILMKDI FHKNNOLAL 181
 Db 125 -----YGNKTKLRELPMNMLQELIGAVRSNNPILCNMTIIMORDIV--ODVELSN 175
 QY 182 TLIDTNRS-RACHPSPMCKSGRCWSESSEDCOSITRTVCAGAGA-RCKGPLPTDCCHQ 239
 Db 176 MSMDOGRHLTGCPKDDPSCPNSSCWGRGENCQTKTICQOCSSRCRGRSPSDCCCHQ 235
 QY 240 CAACTGPKHSDCLALPHNHSICELHCPALVTYNTIDFESMNPREGRYTFGASCYTAC 239
 Db 236 CAACTGPRESDCLVCHFRDEATCKDTCPPLMYNPTTYQMDVNPPEGKYSFGATCYKCC 235
 QY 300 PNYVLTIDVGSCTLVCPLHNOEVTAEQTCRCCKSKCARVCYGLMEHLREYRAVTS 359
 Db 296 PNNVYVTDHSCVRACGPDYEV-EEDVSKCKKCDGCRKVCNGIGIEFKDLSINAT 354
 QY 360 NIOEPAGCKKIFGSLAPLPESEFGDPASNTAPLOPEOLQVEETLEITGYLYISAMPDL 419
 Db 355 NIKHFKYCTAISGDHLIPVAFKGDSTRTPLDRELEIKYKEITGELLQAMPENM 414
 QY 420 PDLVSFQNLQVIRGILINNGASLTLOGLGISWLGSLRSLGSLALIHNNHLCFVHT 479
 Db 415 TDLHAFENLEIRGTHKQGFSLAVWLNTSLTGLSLKISGDVITISGNRLCYANT 474
 QY 480 VPMQDLFRNPHQALHTANPEDECVSGELACHQICARGHMGWGPPTOCVNSCOFJRG 539
 Db 475 IMKKLFGTGNKTKIMNNAEKDOKANHNHNCNPLCSGEGWGEPLDVCSCQVNSGR 534
 QY 540 CYEBCRVUQGLPREYVNAHCLPCHPECOPONGSVTCGPBADCVCACAHYKCPFCVAR 539
 Db 535 CYDKCNILEGEBREVENSECIQCHPECLPQTMNITCTGSRPDCNICAHYVGPCHVKT 534
 QY 600 CPSSGVKPDLSYMPKPFPEDEGACQPCGINCTHSCVDLDXGCAEGRAP-LTSTYSAV 638
 Db 595 CPSSGIMENNTL-VKFPDANNVCHLCHANCTYCAGBGLKGC--QGPBEPKPISTATGI 651
 QY 659 VGILLVVLGVVFGI-LIKRQOKIKRYMYRLLQETELVEPLTPSGAMPNOAMRLIKE 717
 Db 652 VGGLLFIVV-VALGIGLFMRRLQVRRKTLRLQERELVEPLTPSEANQHLRIKE 710
 QY 718 TELRKVXKLGSAGCTYVKGWIPDGENVKIPVAKYLRNTSKXANKELIDEAYVAGV 777
 Db 711 TELFKIKVLGSGAGTIVYKGLMEGEKVKIPVAKELREATSKANKELIDEAYVAVSV 770
 QY 778 GSPYVSRLLGICLTSTVOLVTLMPYGLLDHVENGRIGSODLNMCMQIKKMSYLE 837
 Db 771 DNPVYCRLLGICLTSTVOLVTLMPYGLLDYVHEKDNIGSYLLNMCMQIKKMSYLE 830
 QY 838 DVLVHRDLAARNLVKSPNFKVTTDGLARLIDITEYHADGKVPKIMMALESILR 897
 Db 831 DRLVHRDLAARNLVKTPQVKTITDGLALILGAEBEKYHAEGKVPKIMMALESILR 890
 QY 898 RFTQSDVWSYGVVWELMTGAKPYDGIAPAREIPDLLEKGERLPOPICTIDVYIMTK 957
 Db 891 IYHQSDVWSYGVVWELMTGASRPYDGIAPAREISLLEKGERLPOPICTIDVYIMTK 950
 QY 958 CMWIDSECRPFRELIVEESFMARDPQRFVVIQ-NEDLGASPLDSTFYSLLEDDMDG 1016
 Db 951 CMWIDASRPKFERELILEFSKARDPQRYLVIQDERMHLPSPTDSNFYALMEEDMD 1010
 QY 1017 LVDAEYLVPOQGFPCDDPAGAGAGWHHHRSSSTSGGDLTLGLEBEEBAPSPILA 1076
 Db 1011 VVDAEYLVPOQGF-----NSPST-----SPTPL 1036
 QY 1077 PSAGASDPVPGDLGMAAKGLQSLPHDPSPLQRYSEDPVLPEET--DGVAFLTCS 1134
 Db 1037 SSLANSN-----SSTVACINRNGSCRVKEDAFQRRYSSDPTSVLTEDNDIDFTL----- 1086

QY 1135 PQEYVNPQDVPPSPREGFLPAPAPAGATLERAKTLSPGKNGVWDVAPGAVENP 1194
 Db 1087 PVPEYINQ-SVPRKAGSVQNPVHNGLHF-----AGRDHLION--PHSNASNP 1135
 QY 1195 EYL-TPQGAAPQPPHPPAFSPAFNDLYWQO-----DP-----PERGAPSTF 1237
 Db 1136 EYLNTAQ-----PTCLSGFSDSLMIQKSHQMSLDNPDYQDFFPKAKPNCIF 1186
 QY 1238 KGTPTAENPEYGLDVP 1254
 Db 1187 KG-PTAENAEYLRVAP 1202
 RESULT 3
 Q9EP98 PRELIMINARY, PRT, 1210 AA.
 AC Q9EP98;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Epidermal growth factor receptor isoform 1.
 GN EGFR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVATC;
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
 RA Sinclair C.S., Pearell R.S., Green P.J., Yee D., Lampand A.L.,
 RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
 RA Maillie N.J.;
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative Egr transcripts encoding truncated receptor
 RT isoforms."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
 RA Schehl C., Pearell R.S., Green P.J., Yee D., Lampand A.L.,
 RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
 RA Maillie N.J.;
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative Egr transcripts encoding truncated receptor
 RT isoforms."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF275366; AAC28045.1;
 DR EMBL: AF275365; AAC28045.1; JOINED.
 DR EMBL: AF275367; AAC24386.1;
 DR HSSP: P11362; IFGR.
 DR MGD: MGI:95294; Egfr.
 DR InterPro: IPR000345; CytC heme bind.
 DR InterPro: IPR000494; EGFR_L domain.
 DR InterPro: IPR007119; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR InterPro: IPR01245; Tyr_pkinase.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF01030; Recep_L domain; 2.
 DR PRINTS: PR00109; TRKINASE.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SMO0261; Fu; 5.
 DR SMART: SMO0220; S_TKc; 1.
 DR SMART: SMO0219; TyrKc; 1.
 DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE: PS0011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.

KM ATP-binding: Receptor; Transferase
SQ SEQUENCE 1210 AA; 134840 MM; 62CD021C9DE32E18 CRC64;

Query Match 45.4%; Score 3091; DB 11; Length 1210;
Best Local Similarity 49.1%; Pred. No. 76-226;
Matches 625; Conservative 169; Mismatches 368; Indels 110; Gaps 23;

QY 11 LILALLPGAA--STQCVGTDMKLELPASPTTHLDMLRHLYGQGVVQGNLELTYPNT 68
DB 14 LITLCAAGAGALEEKVVCQGTNRLLTQGTFFDHLSTLQRMNNEVVLGNLETTYQNR 73
QY 69 ASLSFLQDIOEVQVYKANSKEFIGTELQRLRIVRGTOLEFEDNALAVLDGDLNTP 128
DB 74 YDLSFLKTIQEVAGVYLALNTVERIPLENLOIINGNALVENTYALATLS----- 124
QY 129 VMGASPGRLRELQSLTLEILKGVLIQGNPOLCYDTILMKD---PHONQALATLI 184
DB 125 -YGNRTGLRELPMRLQELIGAVFNSNPILCMMDTIQMRDIVQNVFNSMMDL--- 180
QY 185 DTRSRACHPCSPHCKSGSRGWSSEDCSLTRTYCAGGCA-RCKGPLPTDCHEQCA 243
DB 181 -QSHPSCKPCDPSGNSGWSGSENGCKLKIICAQGSHRCKGRSPSCCHNQCAAG 239
QY 244 CTGPKASDCLACHFNHSGICELHCPALVTYNTDFEENPDEGRYTFGASCVTACPY 303
DB 240 CTGPRSDCLVCKPFQDEATCKDTPPLMLYNPITYQMDVNBEGKYSFGATCVKCKP 299
QY 304 LSTDVGSCTLVCPNLHNOETAEQGTORCEKSPCARCYGVGMHELEVAATVANSIO 363
DB 300 VYTDHGSCTVACQCPDYEV-BEDGIRKCKCKCGPRKCKNGIGIEFQDLSITATYKH 358
QY 364 FAGCKKIFGSLAFPEPSFDGPASNTAPLOPBOLOVFTLEITGYLYISAMPDLS 423
DB 359 FKCTAISGDLILPVAFKGDSFTPTPLPRELEILTKVKEITGFLIQAMPDMDLH 418
QY 424 VFQNLQVIRGRLHNAVSLTLOGISMLGLRSRLREGSGALIHNTLCFHTYPM 483
DB 419 AFENLEIRGRKQKQFSLAVAGLNTSLGRSLKEISDGVIIISGRNLCYANTINMK 478
QY 484 QLEFRPHQALLHTANRPEDECVBEGSLACHQLCARGCHGPGFTQCVCNSCPFLRQ 543
DB 479 KLFGTNCKTKIMNRRAEDCAVNVHNCPLSSGSCGMPERDCVSGQNSRRRECEK 538
QY 544 CRYLQGLPREYVNAHQCLPCHPEQPOKGSVTCFPEADQCVACHYKDPFCVAPSG 603
DB 539 CMLBEGPEFEVENSECICQCHPECIPQANNITCTGRPDNICOAHYIDGCHCYCT 598
QY 604 VKPDLSTYMPIMKFPDEBEGACPCPINCSTHSCVDLDDKCPRAEQASPLTSTV 663
DB 599 IMGENNTL-VMKYAADANNVCHLCHANCTYGCAGPGLQCEVWPSGPKIPSIATG 657
QY 664 VVVLGVVFGI-LIKRQOKIRKYTNRRLLQETELVEPLTPSGAMPNOQMKILETEL 722
DB 658 FIVV-VALDIGFMRRRHIVRKETIRLQERELTEPLTPSGEAPNQAHILILETF 716
QY 723 VKVLGSGAGTYYKGIWIPDGENVYKIPAIKYLRENTSPKANKELILEAVYMAV 782
DB 717 IKVLSGAGTYYKGIWIPDGENVYKIPAIKYLRENTSPKANKELILEAVYMAV 776
QY 783 SRLGLCTSTVQVLTQMPYGLLDHVRNRLGSGDLLNMCMQIAKMSYLEDLV 842
DB 777 CRLLGICLTSTVQVLTQMPYGLLDHVRNRLGSGDLLNMCMQIAKMSYLEDLV 836
QY 843 HRDLAARNVLYKSPHNVKITDGLARLLDIDETETHAGGVPIKMAALSILRRTHQ 902
DB 837 HRDLAARNVLYKTPQHVKITDGLAKKLGAEBEKYHAGGVPIKMAALSILRRTHQ 896
QY 903 SDVMSYGVTVWELTFGAKPYDGIAPAREIPDLBEGESLPQPICTIDVYMWKCMID 962
DB 897 SDVMSYGVTVWELTFGSKPYDGIAPADISSILEKGERLPQPICTIDVYMWKCMID 956
QY 963 SECRPRELVESEFRMARDPQRFVLIQ-NEDLGASPLDSTFRSLLEDDDMDGLVD 1021
DB 963 SECRPRELVESEFRMARDPQRFVLIQ-NEDLGASPLDSTFRSLLEDDDMDGLVD 1021

DB 957 ADSRPFRFLIEFSKMARDPQRYVIGQDERMHLPSPTSGNFYALMDEDMEDVYDAD 1016
QY 1022 EYLVPOQGFCDPAPGAGVHHHRSSSTRSGGDLTLGLBSEBEPAPSLAPSEGA 1061
DB 1017 EYLVPOQGF-----NSPST-----SSTPLSLSLA 1044
QY 1082 GSDVPDGLGMAAGLQSLPHDPSLQRISEDPVPLSET--DGVAALTPSGPOPEY 1135
DB 1043 TSN---NSTVACINRNSCRVKEAPLQRYSSDPGTAVTEDNIDDAFL-----PVPEY 1092
QY 1140 VNPQDVPQPPSPREGPLPAARPAQATLERAKTSPGNGVVKDVFAGAVENPEYL-T 1196
DB 1093 VNG-SVPRKBPASVQNPVYHNPPLHP-----ARGDLYQN--PHSNVGNPEYINT 1141
QY 1199 PGGAAPQPHPPAPSPAFDNLVYWDQ-----DP-----PERGAPSTFKGTP 1242
DB 1142 AQ-----PTCLSGSPNSPALMIQSHQMSLDNDPDYQDFEFPKXPKNGIKG-PT 1191
QY 1243 AENPEYLGDPV 1254
DB 1192 AENAEYLRVAP 1203

RESULT 4
ID Q9W6F6 PRELIMINARY; PRT: 1137 AA.
AC Q9W6F6
DT 01-NOV-1999 (TRENBLREL. 12, Created)
DT 01-NOV-1999 (TRENBLREL. 12, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE Receptor tyrosine kinase (Fragment).
GN ERB4.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCB1_TaxID=9031;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=HINDRAIN;
RX MEDLINE=99263203; PubMed=10328884;
RA Dixon M., Lumsden A.;
RT "Distribution of neuregulin-1 (nrg1) and erb4 transcripts in
RT embryonic chick hindbrain."
RL Mol. Cell. Neurosci. 13:237-258(1999).
DR EMBL: AF121963; AAD31764.1; -
DR HSSP: P11362; 1FGK.
DR InterPro: IPR000494; EGRF_L_domain.
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR002174; Tyrosin-1ike.
DR InterPro: IPR001368; TNFR_c6.
DR InterPro: IPR001245; Tyr_Pkinase.
DR InterPro: IPR004019; YLP_motif.
DR Pfam: PF00757; Furin-1ike; 1.
DR Pfam: PF00069; Pkinase; 1.
DR Pfam: PF01030; Recep_L_domain; 1.
DR Pfam: PF02757; YLP; 2.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_Pkinase; 1.
DR SMART: SM00261; FU; 3_Pkinase; 1.
DR SMART: SM00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS00652; TNFR_NGFR_1; UNKNOWN_1.
KW kinase; Tyrosine-protein kinase.
FT NON TER
SQ SEQUENCE 1137 AA; 127927 MW; 4DE16436F7DC84F CRC64;
Query Match 40.1%; Score 2729.5; DB 13; Length 1137;
Best Local Similarity 47.3%; Pred. No. 26-198;
Matches 539; Conservative 169; Mismatches 354; Indels 77; Gaps 24;

```

QY 161 LCYODTILMKDIFHKNOALTLIDTNSRACHPCSPMKGSRGCESSDDCQSLTRTVG 220
DB 3 LCPADTTHMODIRNPMASNFLLVPTNGSSGCRGCKSCGT-RCMGPTENHCQTLTKVVC 61
QY 221 AGGC-ARCKGPLPTDCCHQCAAGCTGPRKSDCLACIHNHSGITELHCPALVTYNTDTF 279
DB 62 AECCDGRCTGFPYVSDCHRECAAGCGEPDTCFCAMFNDSGACVTCQCPOTFVYNPTTF 121
QY 280 ESNPNEGRTYFGASCVTACFPYVLTVDVSGCTLVCPHNOEVTAEQTCCEKSKPCA 339
DB 122 QLEHNNAKYTAGAFCKKCPHNHV-VDSGCVRAKSPSKHEV-EENIKMKCKCTDIDCP 179
QY 340 RVCYGLGMEHLREVRVTSANIQEPAGCKKITFGSLAFPESDGDPASNTPLPDEQICV 399
DB 180 KACDGIQTSLSVAQIVDSSNIDKFINCKTNGNLIPLVTGIHGDPPYHTIAINPEKINI 239
QY 400 FETLEITGLYISAMPDLPDLVSFQNLQVIRGRILHNNGAYSITLQGLGSMGLSLR 459
DB 240 FQTVREITGYLNIQSPENMTDFRVFSNLVTIGBALYSGLSLILKQGITSLQFQSLK 299
QY 460 ELGSGIALIHNTHLCFVHTVPWDLFRNPHQALHTANPEDECVSGELACHQLCARGH 519
DB 300 QISAGNIYITDNSNICYHTVMTSLFSTPSQKVIHNKKAENCTADGMVCNELCSSDG 359
QY 520 CWPGRPTOCVNSOPFLRGOECVBEGRVLCGLPREVYNAHCLPCHPEQCP-QNSVYCFG 578
DB 360 CWPGRPDQCLSKRPIRGRTCIESCNIYDGEFRFPAKNSVCMECDPQCEKMDNMTTCYG 439
QY 579 PEADOCVACAHYKDPFVCARCPSPGVKBDLSYMPIWKEPDEGACOPCPINCTSCVDLD 638
DB 420 PGPDHCTKCFHFKDPNCEKCPDGLQGANSF--IFKYADEDRCHPHNCTGCGGCPA 477
QY 639 DKGC-----PABGRASPLTISYSAV-GLLVVYGVFGLIKRQOKIRKYT 666
DB 478 SHDCIYPTWROSTLPOHAR-TPL--IAGVIGGFIIVIMGLFAYVYKRSK-K-KRA 533
QY 687 MRLLQETELVEPLTPSGAMPNOAOMRIKETEELKRVYLSGAFGVYGIWIPDGENV 746
DB 534 LRRL-ETELVEPLTPSGTAPNOQLRIKTELKRVYLSGAFGVYGIWIPDEETV 592
QY 747 KIPVALIKVLRNTSKANKELDEYVYVAGSGSPYSRLGICLTSTVOLVTOIMPYGL 806
DB 593 KIPVALIKILNETTGKAVEMDEKLIVASMDHPLVRLGLVCISPTIQLVTOIMPYGL 652
QY 807 LDHVENRGRIGSODLIMCMQIAKMSYEDVRLVHRDLAARVVLKSNHYKIDPGL 866
DB 653 LDYVHEHKNIGSOLLIMVCQIAKMYLEERLVRHDLAARVVLKSNHYKIDPGL 712
QY 867 ARLLDIDETVHADGKVPKIMWALESTLRRFTHOSDWSYGVTVMLMTFGAKPYDGI 926
DB 713 ARLEGEDEKEYNADCGKPIKIMWALECIHYRKFTHOSDWSYGVTVMLMTFGKPYDGI 772
QY 927 PAREIPDLLEKGERLPOPICTIVYIMVYKMWIDSECPREPLVSEFSRAARPORF 986
DB 773 PIREIPDLLEKGERLPOPICTIVYIMVYKMWIDDSPKFELAEFSRAARPORF 832
QY 987 VVIQNEQ-LGPASPLDSTFYRSLLEDDMDGLVDAEYVLPQOGFFCPDPAAGAGVMH 1045
DB 833 LVIOGDDRMKLPSPNDKSPFQNLDEEDLEMDMAEYLV-CAFNIPPIYISRRIRDS 891
QY 1046 RHSSSTPSGGGDLTLGLEPSEEARPS--PLAP-SGASADVFDGDLGMAKGIQSLP 1102
DB 892 NRNOFTVRDGGYAAEQGV-PMYPAPGCIIPFAVAVAGATALEIFEDTCNGTLRKQVATL 950
QY 1103 THDSPLOQVSEDPVLPF-----ETDGYVAPLTCSPQPEYVQAPVPAPSPFREG 1155
DB 951 AKEDSSTQVYADPTVFLPERVINGELDEDDGYMTPRMDRKTYLPLVLENPVSRRKKG 1010
QY 1156 PLPAA-RPAGATLERAKTILSPGKGVYKVF-----AFGAIVENPEYLTQGGAAQ 1206
DB 1011 DLQAVDNPBYHN-----APNGQPKADEYVNEPLVLTNTFANTLENAEYL-----K 1055
QY 1207 PHPPAFSPAFDNLVYWDOPPERGA--PSTFKGPT-----AENPEYL 1249

```

```

DB 1056 NNUPREAKKAFNPDPYMNHSLPRSTLGHPTDQIYESTKYFYKONGRIPIVAENPEYL 1114

RESULT 5
Q9YH40
ID Q9YH40 PRELIMINARY; PRT: 1165 AA.
AC Q9YH40
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Receptor tyrosine kinase proto-oncogene.
GN XMRK.
OS Xiphophorus xiphidium.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorphi; Atherinomorphi;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_TaxID=8086;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIO PURIFICATION;
RX MEDLINE=98241172; PubMed=9582016;
RA Dimitrijevic N., Minkler C., Wellbrock C., Gomez A., Duschl J.,
RA Altschmid J., Scharf M.,
RT "Activation of the Xmrk proto-oncogene of Xiphophorus by
RN Oncogene 16:1681-1690(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RIO PURIFICATION;
RA Scharf M.,
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: U53471; AD010500.2; -.
DR HSSP: P1362; IPGK.
DR InterPro: IPR000145; Cytochrome bind.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001899; Gram pos anchor.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00057; Furin-like; 1.
DR Pfam: PF00069; Pkinase; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00261; Fu; 3.
DR SMART: SM00219; Tyrc; 1.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00011; PROTEIN_KINASE_TYR; 2.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1165 AA; 129614 MW; 7FEB38D8771A74E CAC64;

Query Match 39.7%; Score 2703; DB 13; Length 1165;
Beet Local Similarity 45.3%; Pred. No. 2,le-196;
Matches 576; Conservative 163; Mismatches 393; Indels 140; Gaps 29;
QY 1 MELAALCRWGILLALLPPG-AAST---QVCTGDMKRLRPAPEFHLMRLRYGCCOV 55
DB 4 LELEL-----LHLLLSIGRCCSTDPKRVCGSTSNQMTV--LDNHYYLKKMKWSSGCV 56
QY 56 VQGNLELTYPTNASUSFLDQIGVOOYIKANSKFIGITELQRLIVRGSTQLFEDNYALA 115
DB 57 VLENLEITYQENQDLSFLQSIGEVGYVLIANNVSTIPLVVLRIROGNLYEGNFTLL 116
QY 116 VLDNGDPLNNTFTVYASPGGLRELDRLSTELTKKGVLIQRNPQLCYDPTLWKDIFHK 175
DB 117 VMSYVQK-APSSP--DYQVQKQLQSLNLTSLTSGVAVSNPCLCNVETINMWDIVDK 173
QY 176 NNOLATLIDTNSRACHPCSPMKGSRGCESSDDCQSLTRTVGAGG-ARCKGPLPTD 234

```

```

Db      174 TSNPTMNLIPHAPEKCCQKDPGCVNGSMAPRGHCQKFTLLLAECQCNRCRGPID 233
QY      235 CCHQCAAGCTGPKHSDCLAHFNHSGICELHCPALVTYNTDTESMPNEGRTFGAS 294
Db      234 CCHNHCAAGCTGPRATDCLACRDNDDGTCKDCTPPKIIDIVSHQVVDNPNIKYTCGA 293
QY      295 CTTACCPYNTLSTDVSGCTLVCPHNOETAEQDQRCCKSCSPCARVCGYGMHELRVR 354
Db      294 CVKECPSNVYVE-GACVRCSGAGMLEVD-ENGKSKCKPCDGVCPKVDGIGISLSTI 351
QY      355 AVTSANIOEFAGCKKFFGSLAFPESEFDGDPASNTAPLOEQLVFETLEITGLYISA 414
Db      352 AVNSTIGSFNSNCTKINGDILNRSFEEDPHYKIGPMDPEHMLMTLVKEITGLVIMW 411
QY      415 WPDLSPLDLSVFOQLQYIRGRIILHNKAYS-LTQIGISWLSGIRSLREISGALIHNTH 473
Db      412 WPMNTSLSVFQMLETIRGRTTFSRGFSFVVOVSHLQMLGSLSKREVSANVILKNTPO 471
QY      474 LCFVHTVPMQDLFRNDHQAHLHTANPEDECVGEGLACHOLGARGHCWGPPTCCVNSQ 513
Db      472 LRYASTINMRRLFRSHDQSI EVDART-----ENQTCNNESESDGCGPPTMVCGLH 524
QY      534 FLRGQCEVEECRYLQGLPREYVNAHCLFCHPECCPQNGSVTCFPEPADQCAAHYKP 593
Db      525 VDRGRCVVASCNLLOEPRBAQVGRVQCHOECVQDLSLTCYGPGRANSGKCAHFQDG 584
QY      594 PFCVACPSGVKPDLSYMPIMKFPDEEGACQPCPINTCHSCVDLDDKCPAEGASPLTS 653
Db      585 PCITPCPCPHGMLGDSGTL-IMKADMGCCQCHONCTQCGSGPGLSCGRD-IVSHSL 642
QY      654 IYSAVVGILLVVLGVVFSILIKRQOKIRKYMRRLLQETELVEPLTPSGAMPNOAMR 713
Db      643 AVGLVSGLLITVVALIIVLLRRRIK-RKRTIRLLQEKELVEPLTPSGAPQAPQAFIR 701
QY      714 ILKETELRKVYLGSGAFGVYKGIWIPGENVVKI PVALKYLRNTSKAKELIDEXYV 773
Db      702 ILKETEFKDRVLSGAFGVYKGLNIPGENIRIPVAILKYLRETSKRVQEVLDXYV 761
QY      774 MAGVSPVYSRLIGLICLTSTVOLVTOLMFGCLDHVSENRRGLSODILMCMQIAKM 813
Db      762 MASVDHPHYCRLLIGLICLTSAVQLVTOLMFGCLLDYVQHOGRICGQMLMWCQIAKM 821
QY      834 SYIEDRLVHRDLAANVLVKSPPNHYKIDPGLAALLDIDETEHYADGKVPIMMALES 893
Db      822 NYLEERHLVHRDLAANVLVKNPNHYKIDPGLSKLLADEKEXYAHGKXVPIKMMALES 881
QY      894 ILRRFTHQSDVMSYGVTVWEIMTFGAKFYDGIIPAREIPDLLEKGERLPQPICTIDYV 953
Db      882 ILQMTYTHQSDVMSYGVTVWEIMTFGSKYDGIIPAKETASVLENERLPPQPICTIEYV 941
QY      954 IMYKCMWIDSECPRRRELIVSEFSKMAHPQPFVVIQNEIDGPAFLDSTYRSLLBEDD 1013
Db      942 IILKCMWIDPSSPRRELIVGFSQMAHPSPRYLVIQ--NLPSPSRRRFFSLLSSDD 998
QY      1014 MGDLYAESEYLVQQCFPCPPAPAGAGVHHRHSSSTRSGGDLUTGLPSEBEAPRS 1073
Db      999 --DIVDADEYLL-----RYKRI-N-RQSS-----E 1019
QY      1074 PLAPSEGASDVFDGDLGMAKAGUOLPETHDPSLPQRYSEDPV-PLPSETDGYVAFLT 1132
Db      1020 PCIPPNGH-----FVRENSIALRYISDPTQNALMEKLDGH----- 1054
QY      1133 CSPPEYVNOPIVRPQ-----PSPRE-----GPLP-AARPAATILERAKTISPGNG 1179
Db      1055 -----EYVNOPOSSETSRLSDIYNRYEDLTGKWPVSLSGEATNFSREYVNTQNS 1109
QY      1180 VVKDVAFAGGAVENTPEYLTPOGGAAPQPHPPAPSPAFDNLVYMDQDPERGAPSPSTFK 1239
Db      1110 L--PLVSSGSMDDPDY--QAG-----YQAAF-----LPQGALTGKGMF 1144
QY      1240 TPTAENPEYIGL 1251

```

```

Db      1145 LPAENPEYIGL 1156

RESULT 6
P79754
ID P79754 PRELIMINARY; PRT: 1326 AA.
AC P79754;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ekb3.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99177347; PubMed=10077531;
RA Gellner K., Brenner S.;
RT "Analysis of 148 kb of genomic DNA around the wrt1 locus of Fugu
RT rubripes.";
RL Genome Res. 9:251-258 (1999).
DR EMBL; AF056116; AAC34391.1; -.
DR HSSP; P11362; 1FCR.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; kinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Prodom; PD000001; Euk_kinase; 1.
DR SMART; SM00261; RU_3.
DR SMART; SM00219; TYKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KM ATP-binding; Transferase.
SQ
SEQUENCE 1326 AA; 148613 MW; A33303258B647E9 CRC64;

Query Match 33.4%; Score 2275; DB 13; Length 1328;
Best Local Similarity 40.2%; Pred. No. 9,1e-164;
Matches 517; Conservative 152; Mismatches 427; Indels 190; Gaps 31;

9 MGLTALLPP--GAASTQ---VCTGDMKRLTPASPTHLMRLHYQGGVQVQGNLEL 62
4 MLLIMCVASRLRAASSQTEAVCGCTQNGLSSTGQENQYNLNKDRYKGCETIMGLLEI 63
63 TYLPTNASLSFLQDIQEVQOYIKANSKFIGITELQRLRIVRGTQLFEDNYVALVLDNGDP 122
64 TQIESMWPFSFKLTIREYTGYYLLIMNHFOEIPLGQLRIVIGNSLYERRFALSVFLN--- 120
QY 123 LNNTPVYTAGSGGLREQLRSITELILGVLQIRNPOLQVQDITIMWDIFHKNNQALALT 182
Db 121 ----YPKDC--PSGLNQLGLMNLTEILLGVOIINNKYLIRGPMVYMRDII--RNNDAPIE 173
QY 183 LIDNRSRACHPCSPCKSGSRWCSSSDCSLTRYVCAGGC-ARCKGPLPTDCHBOCA 241
Db 174 IQFNGBRGVCH--KSC-GNYCWMGPKQDQCIILFVTCVAPCCNDRCFQTSBRDCHIECA 229
QY 242 AGCTGPKHSDCLAHFNHSGICELHCPALVTYNTDTESMPNEGRTFGASCVTAQCPY 301
Db 230 AGCKGGLDIDCFACKLFSDSGAVCPQCGQTLLYNNKQTFQMETNNAKYOGASICVSCPT 289
QY 302 NYLSTDVSGCTLVCPHNOETAEQDQRCCKSKPCARVCGYGMHELRVAVVTSAN 360
Db 290 HRV-VDGSCVAVCPDPMVEV--ERGSQRQCELSGCLCPKXCETGAE---QRTVDSSN 343
QY 361 IQEPAGCKKIFGSLAFLESEFDGDPASNTAPLOEQLVFETLEITGLYISAMPUSLP 420
Db 344 IDSFNCTRIQSLHFLVTGILGDFKXVPPLDKCUVEFRVAREITDIINTIOSWPKELN 403

```

```

QY 421 DLSVFQNLQVIRGRILHNGAYSLLTLOGIISWLGLRSRLRELSGLALIHNTHLCPVHTV 480
DB 404 DLSVFSSLTTCORSLFKRFSILWMRIPTLTSLGRSLREISDSGVISQNAHLCHHTV 463
QY 481 PWDQLFRNH-QALLHTANRPEDECCYGBELACHOUICAHGCKGPEPTCCVNCOSQFLRGQE 539
DB 464 NMTQLEFRGRVRANSINSRPMACVADRVCDPLCSGSGCWPDPDCLSCRYSRHGT 523
QY 540 CVEECRVLOGLPREYVNAH-CLPCHPECOPONGSVTCFGEPAADCVACAHYKDPFCVA 598
DB 524 CVAGCHFNSGIRPREFGLNGVACVACHPEKPTGTASCTGAGDECACTFRGPTCMS 583
QY 599 RCPSSGVKPLSTWPIWKFPDEBGAQCPFCINCHSCVLDKKGCPAECRASPLRTISAV 658
DB 584 SCPAGVN-DGEGKLIKFEPNREGHCEPHONCTQCGSPGLNDC--LEAARLTSSGQI 639
QY 659 VGILLVVLGVVF-----GILIKRQOQKIRKYMRLLOTELEVEPLTSSGAMPQAO 711
DB 640 TGIALGVPAGLIFCLVLPFLGMLYRGLAIRKRAMRYLESSESFEPLGF-GEKGTKVH 698
QY 712 MRLKETELRKVKVIGSGAFGVYKGIWIPGGENYKIPVAIKVLRENTSPKANKELIDEA 771
DB 699 ARLKPSDLRKIKPLGSGVFTVSGKFWPEGETYKIPVALKTIQDSSGRQTFTEIDHL 758
QY 772 YVMAGVSPYVRILGICLTSTVQVLTQLMRYCCLDHYRENRGLSODILMKCQAIK 831
DB 759 LSMGSLDHPYIVRLGICPGTCLQVLTSSHGLLEIRHKTSLDORLMMCVQAIK 818
QY 832 GMSYLEDELVLRDLAANVAVKSPNHVKTDFGLARLLDIDETEHADGKVPKIMMAL 891
DB 819 GMYLEBRVHVHKNLAANILLKNDYQVQISDYADLLPYDDKYYVSEKRTPIKMAL 878
QY 892 ESILARRFTHSDVWSYVTVWELMTFGAKPYDGIIPAREIDLEKEBRLPPOPTCTIDV 951
DB 879 ESILARRFTHSDVWSYVTVWELMTFGAKPYDGIIPAREIDLEKEBRLPPOPTCTIDV 938
QY 952 YWIMVCKWVIDSECPREFRELSEFSRMAPRPOFVYQNDLCPAPBLDSTFRSLLED 1011
DB 939 YWIMVCKWVIDSECPREFRELSEFSRMAPRPOFVYQNDLCPAPBLDSTFRSLLED 982
QY 1012 DDMGDLVDAEYLVQOGFPCPDPAAGVHHRRSSSTRSGGDLTLGLEPSEEAR 1071
DB 983 SGMGEFL-----RRGSR--GLLEADLEDEDE-- 1008
QY 1072 RSPILAPSEGAGSDVDDGLMG--AAKGLQSLPTHPSPIQ-----RYSEDPVY 1118
DB 1009 -----GLGDRFATPSLDOPSPSWSTSPQINSYVMTQLRD----- 1044
QY 1119 PLPSETDGVAPLTCSPQ--EYVQ-----PDVPPQSPSPREGPL--PAAR 1161
DB 1045 -FAVSGGHIGYLPSPSPVDITIRQLWYQSRSLSSVRLPDRSAFRSSSEALCEDGAQ 1103
QY 1162 PAGATLEBAKTLSSGKHGVADVAFGAVENPEYLLPQCGAAPQHPHPPAFSPAFDNLV 1221
DB 1104 CAGIFRVA-----FSSERGN-----PQGG----- 1122
QY 1222 YWDQDPERGAPSTFKGTPTAENPE 1247
DB 1123 --OORKLSTASSPSFKTWADEDE 1146

```

RESULT 7

```

Q9BIH9 PRELIMINARY; PR7; 1433 AA.
AC Q9BIH9;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;

```

```

OC Anopheles.
OX NCBI_TaxId=7165;
RN 1
SEQUENCE FROM N.A.
RC STRAIN=GA.
RA Lycett G.U.;
RT "Cloning, expression and localisation of the Anopheles gambiae
epidermal growth factor receptor."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301655; CAC35008.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000345; GyrC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002280; Set-ctr_Pkinase.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; kinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON_TER.
SQ
SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;
Query Match 29.5%; Score 2011.5; DB 5; Length 1433;
Best Local Similarity 32.3%; Pred. No. 1,1e-143;
Matches 467; Conservative 197; Mismatches 388; Indels 385; Gaps 36;

```

```

QY 549 GLPREY-VNARHCLPHNPECOQPNQSVTCFGEADQVCAAHYKQPPCVARCP----- 601
DB 507 SLRLYVSVDKTCGDHCECKD-----FCYGRPNQCGSCMMVXKGRFCVACPTTKAM 561
QY 602 -----SGVFPDLSTYMPIWKFPD----- 618
DB 562 NGCINCINCHTKVCGRCRPRDTIAPDGCISCKKAIIGSDALIEHCLMKKSDCPDGYSDYVL 621
QY 619 -EEG----- 621
DB 622 QEEGFLKQLSGKAVCRKCHPCKCTGYGFHFOFCOECTGYKKGCEDECPDIFYANEE 681
QY 622 --ACQPCPINCT-----HSCVDL-----DD-----XGCPAEQ----- 646
DB 682 TRICLPHOECRCCHGLGDDHCECRNLKLFEGDPYDNATFTFCVNCNCASHYKCFPOEA 741
QY 647 -----RASPLTISVAVVGLLVVVGVI-----LKRQOKIRKXTM 687
DB 742 GKIGPYCSADSMOSGRIEFPOTGVKIVMSVVALILLCVFGIAFVLESRHKKKDAVKM 801
QY 688 RRLLOETELVEPLTPGAMPNCAOMRILKETELRKVKVLGSGAFGYVKGIKIPGCVNK 747
DB 802 TMALAGCEDESEPLRPSVGNPNLTKLRIEASIRKGVLGKGAFFRVRFGVAMPGESEVK 861
QY 748 IYVAIKVIRENTSPKANKELDEAYVAVGSPYVSRLLGICLSTVQLTQLMFYGCLL 807
DB 862 IYVAIKVIRENTSPKANKELDEAYVAVGSPYVSRLLGICLSTVQLTQLMFYGCLL 921
QY 808 DHYRENRGRIGSODLNNMCQIAKMSYLEDVRLVHRDLAANVLKSPNRYKITDEFLA 867
DB 922 DYVRNNKDKIGSALLNWSQIARGLAAYEERLVHRDLAANVLVQPSCKRIVFELA 981
QY 868 RLIDIDETEHADGKVPIKMMALLESILRRRFTHOSDVSYGVTWELMTFGAKPYDIP 927
DB 982 KLIDFDSDEYRAGGMPIKMLALIECIRHVFSTKSDVAFGITITWELTYARAYENVP 1041
QY 928 AREPDLLEKEXELPQPICTIDVYIMVKCMIDSECFRRRELVSFESMARDPQFV 987
DB 1042 AKDVELIEIGHKLPPQDICSLLVYICILSCVWLADAPTRKQLAETFAEXARDPGRYL 1101
QY 988 VIONEDIGPASPLDSTFYRSILEDDPMGLV----- 1018
DB 1102 MT-----PQDKMRRLPSYTNODEKDLIRTLARVAMAAAAAAGASNVDPSTIA 1152
QY 1019 DAEYVLVQCGFCPPAPAGAGMVHHRHRSSTYSGGDLTLGLEPSEEARPS----- 1073
DB 1153 ETDVYLQPKTRPSIMLPGBSA-----VEPS-DEMPKSLRYCK 1188
QY 1074 -PLAP--SEGASDVFDDGLGWAAGKGLQSLPTHDPSPFLQRYSDPTVPLPSETDGYVA 1129
DB 1189 DPLKPDDETGHGKEY-----GVGGR-----LNDPLDEDDYLM 1222
QY 1130 PLTCSPOFEYVNPQVPRPQPSREGRPLPARPAGATLERAKTLSPKNGVVKDYFAFGG 1189
DB 1223 P-TQSGQ-----NQS-----TPG-----YMDLIGVPA 1243
QY 1190 AVENPEYL-----TPQGAAPQHPHPAPAFDFVLYWMDPFBKGPSPSTFGKT 1240
DB 1244 SVDPPEYLMGSTALAGLQAGSWG--PHTP-----PPTPNGM 1280
QY 1241 PTAENPE 1247
DB 1281 PTHQSG 1287

```

RESULT 8
Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Herstatin.

```

CN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Byars A.J., Hennen W.D.;
RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF177761; AAD56009.2;
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FECLBB347E2D030C CRC64;

```

Query Match 26.4%; Score 1797.5; DB 4; Length 419;
Best local Similarity 86.3%; Pred. No. 3,3e-128;
Matches 340; Conservative 7; Mismatches 40; Indels 7; Gaps 2

```

QY 1 METALCRNGLLALLPRAASTQVCTGDMKRLRPAPEHTLDMRLHYGQCVVQGNL 60
DB 1 METALCRNGLLALLPRAASTQVCTGDMKRLRPAPEHTLDMRLHYGQCVVQGNL 60
QY 61 ELTYLPTNASTFLQDIPQVQYKANSKFISITELQSLRIYRGQLPEDNYALAVLNG 120
DB 61 ELTYLPTNASTFLQDIPQVQYKANSKFISITELQSLRIYRGQLPEDNYALAVLNG 120
QY 121 DPLNNTPTVYASPGGLRELQRLSLTEILKGVLLIQNPQLCYQDTILMKDIFKNNGLA 180
DB 121 DPLNNTPTVYASPGGLRELQRLSLTEILKGVLLIQNPQLCYQDTILMKDIFKNNGLA 180
QY 181 LTLIDNBRACHPQSPMKSGSRGSESESDQSILTRVCAAGCARCGPLPTDCHEQC 240
DB 181 LTLIDNBRACHPQSPMKSGSRGSESESDQSILTRVCAAGCARCGPLPTDCHEQC 240
QY 241 AAGCTGPKSDCLACLFHNSGICELHPALVTYNTDIFESKPNBGRYTTGASCVTACP 300
DB 241 AAGCTGPKSDCLACLFHNSGICELHPALVTYNTDIFESKPNBGRYTTGASCVTACP 300
QY 301 YNLTSDVQSCILVCPHNOEYTAEDTGRCCKSKPCARVYGLGMEHLAEVAVTSAN 360
DB 301 YNLTSDVQSCILVCPHNOEYTAEDTGRCCKSKPCARVYGLGMEHLAEVAVTSAN 360
QY 361 IOFAGCKKIFGSLAFLPESFGDASNTAPLP 394
DB 356 LRMQPG--PAHFVLSLRPSWDLVAFSLPLAP 387

```

RESULT 9
Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/Genbank/DBJ databases.

DR EMBL; BC027080; AAH27080.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match 25.5%; Score 1739; DB 11; Length 367;
 Best Local Similarity 88.0%; Pred. No. 7, 5e-124;
 Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 889 MALESILRRRFTQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLERGERLPQPICT 948
 DB 1 MALESILRRRFTQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLERGERLPQPICT 60
 QY 949 IDVYIMVKKCMIMDSGCRPRFRELVSFGRMARDPQRFVVIQNEDELGPASPLDSTFYASL 1008
 DB 61 IDVYIMVKKCMIMDSGCRPRFRELVSFGRMARDPQRFVVIQNEDELGPASPLDSTFYASL 120
 QY 1009 LEDDDMGDLVDAEEYLVPOGFFCPDPAPGAGMWHRRSSSTSGGGDLTLGIEPSEE 1068
 DB 121 LEDDDMGDLVDAEEYLVPOGFFCPDPAPGAGMWHRRSSSTSGGGDLTLGIEPSEE 180
 QY 1069 EAPRSPLAPSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOQYSEDPTVP.LPSETDGYV 1128
 DB 181 EAPRSPLAPSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOQYSEDPTVP.LPSETDGYV 240
 QY 1129 AETCSPOPEYVNOQPVRRPQPPSPREGPLPARPAGATLERAKTISFGKNGVYKVFAPFG 1168
 DB 241 APLACSPQPEYVNOQPVRRPQPPSPREGPLPARPAGATLERAKTISFGKNGVYKVFAPFG 300
 QY 1189 GAVENEYLTPOGGAPOPFPAPFSPAFDNLVYMDQPEERGAPPTFFKGPTEANEY 1248
 DB 301 GAVENEYLTPOGGAPOPFPAPFSPAFDNLVYMDQPEERGAPPTFFKGPTEANEY 360
 QY 1249 LGLDVYV 1255
 DB 361 LGLDVYV 367

RESULT 10
 086712 PRELIMINARY; PRT; 729 AA.
 AC 086712;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Polyprotein.
 GN POLYPROTEIN.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
 OX NCBI_TaxID=11950;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; PubMed=8152791;
 RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
 RA Johnson A., Beug H.;
 RT Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities.";
 RL Oncogene 9:1307-1320(1994).
 DR EMBL; S69372; AAC60725.1; -
 DR HSSP; P03322; IAS6.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR004028; Retro_M.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF02813; Retro_M; 1.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00219; TyrKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EFELD63 CMC64;

Query Match 25.3%; Score 1720; DB 15; Length 729;
 Best Local Similarity 54.8%; Pred. No. 5, 9e-122;
 Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15

QY 569 PQNGSVTCFQPEADQCAAHYKDPFCVAPSGVDPDLSYPIWKFPDEGACQPCPI 628
 DB 141 PEETATPRTGP--DHCMKCAHFIDGPHCVKACPVAGLGENDTL-VWKYADANAACQLGHP 197
 QY 629 NCHSCVDDDDKCCAPAEQASPLTISVSAV-V-GILLVVLGVVFGILLRQOKTRKTYM 687
 DB 198 NCTRGCKPGECCP---NGSKTPSTIAAGVAGLCLVVGIGIGLYLRR-HIVRKTL 253
 QY 688 RLLOETELVEPLTPSGAMPNOQMRLKETELRKVKVLGSAFGVYKGIWIPGENVK 747
 DB 254 RLLOETELVEPLTPSGAMPNOQMRLKETELRKVKVLGSAFGVYKGIWIPGENVK 313
 QY 748 IPAIVKLRNLTSPKANKELDEAYVAGVSPYVRLIGLSTVQLTQMLPFGCLL 807
 DB 314 IPAIVKLRNLTSPKANKELDEAYVAGVSPYVRLIGLSTVQLTQMLPFGCLL 373
 QY 808 DHVRENRGRLSQDLTLMCMQIAKGSYLEDVRLVHRDLAARNVLKSPNHVKTDFGLA 867
 DB 374 DYIREKDNIGSOYLLNMCVQIAKGNVLEERLVHRDLAARNVLKTPQHVKITDFGLA 433
 QY 868 RLIDIDETVHADGKVPKXMALESILRRRFTQSDVWSYGVTVWELMTFGAKPYDGI 927
 DB 434 KILGADEKEYHAEGKVPIKXMALESILRRRFTQSDVWSYGVTVWELMTFGAKPYDGI 493
 QY 928 ARRIPLLEKGERLPQPICTIDVYIMVKKCMIMDSGCRPRFRELVSFGRMARDPQRFV 987
 DB 494 ASEISVLEKGERLPQPICTIDVYIMVKKCMIMDSGCRPRFRELVSFGRMARDPQRFV 553
 QY 988 VIQ-NEDLGAPASPLDSTFYRSLLDDDMGDLVDAEEYLVPOGFFCPDPAPGAGMWHR 1041
 DB 554 VIQDERMHLPSPTDSKFRITMEEDMEDIDADAEYLVPHQGF----- 598
 QY 1047 HRSSSTRSGGDLTLGLEPSEEARPSL-----APSGAGSDVFDGDLGMAKGLQSL 1101
 DB 599 -NSPST-----SRTPLSLSLSTSNNSATNCID-----RNGGCH 631
 QY 1102 PTHDPSPLOQYSEDPTVP.LPSET--DGYVAPLTCPOPEYVNOQPVRRPQPPSPREGPLPA 1155
 DB 632 PTHDPSPLOQYSEDPTVP.LPSET--DGYVAPLTCPOPEYVNOQPVRRPQPPSPREGPLPA 1155
 QY 1160 ARPAGATLERAKTISFGKNGVYKVF-----AFGGAVENPEYL 1197
 DB 676 -----TAWVQNVYNNISLTAISKLPWDSRYQNSHSTAVDNPEYL 715

RESULT 11
 086714 PRELIMINARY; PRT; 567 AA.
 AC 086714;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE v-erbB protein (Fragment).
 GN v-ERBB.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
 OX NCBI_TaxID=11950;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; PubMed=8152791;
 RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
 RA Johnson A., Beug H.;
 RT Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities.";
 RL Oncogene 9:1307-1320(1994).
 DR EMBL; S69372; AAC60727.1; -
 DR HSSP; P11362; IFGK.
 DR InterPro; IPR000719; Euk_pkinase.

DR InterPro: IPR001245; Tyr_kinase.
 DR Pfam: PF00069; kinase.1
 DR PRINTS: PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk_kinase.1.
 DR SMART: SM00219; TyrcKc.1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP.1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM.1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR.1.
 DR Tyrosine-protein kinase.
 KW NON TER
 FT
 SQ SEQUENCE 567 AA; 63390 MW; C6D9CBAVAD725E1 CRC64;

Query Match 25.2%; Score 1718; DB 15; Length 567;
 Best Local Similarity 55.4%; Pred. No. 5.7e-122;
 Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 578 GREADQVCAAHYKDPPEVCVARGCSGVKPDLSVPIKPFDEGACQCPINCTHSCVDL 637
 DB 1 GP--DHCKKCAHFIDGPHCVACAFAGLGENDTL-VKYNADANAVCQLCHPNCRCCKGP 57
 QY 638 DDKGPAPQASPLTSIVSAVY-GILLVVLGVVFGILIKRQOKIRKYMRLDQETEL 696
 DB 58 GLEGCP--NGSKPTSIAGVYVGLCLVVGIGLYLRR-HYRKRTIRLLQEREL 113
 QY 697 VEPLTPSGAMPNOMRLKETELRKVKVYGSAGFTVYKGIWIPGENVKIPVALKVLK 756
 DB 114 VEPLTPSGEAPNQHRLILKETEFKVKVYGSAGFTVYKGIWIPGEKVKIPVA-KEEL 173
 QY 757 ENTSPKANKELIDEAAYMAGVSPYVSRLLGICLTSTVQVLTQMLVYGCILDHVRNRR 816
 DB 174 EATSPKANKELIDEAAYMASVDNPRVCRLLGICLTSTVQVLTQMLVYGCILDHVRNRR 233
 QY 817 LGSODLLNMCQIAKNGSYLEDVRLVARDLAARNVLYKSPNHVYKIDFGIARLLDDETE 876
 DB 234 IGSQYLLNMCQVQIAKNGSYLEERLVARDLAARNVLYKTPQHVKIDFGIARLLDDETE 293
 QY 877 YHADGKVPFKMMALESILRRFTHQSDWVYGVYVWELMTFAKRYDGIIPARELPDLE 936
 DB 294 YHAEAGKVPFKMMALESILRRFTHQSDWVYGVYVWELMTFAKRYDGIIPARELPDLE 353
 QY 937 KGERLPQPPICITIDVYIMVYKCMWIDSECRPFRELVESEFRMARDPQRFVYIQ-NEIDG 995
 DB 354 KGERLPQPPICITIDVYIMVYKCMWIDSECRPFRELVESEFRMARDPQRFVYIQ-NEIDG 413
 QY 996 PASPLDSTFYRSLLEDMDGLVDAAEYLYLPQGFPCPDPAAGAGMVAHRRSSSTRSG 1055
 DB 414 LPSPTDSKFRYRTIMEEDMEDIVDADEYLVPHQGF-NSPST- 454
 QY 1056 GGDLLTGLPESEEPSPPL-APSEGAGSDVFDGDLGMAKGLQSLPTHPSPQL 1110
 DB 455 -----SRPLLSLSATSNNSATNCID-----RNGQGHVYREDSEVQ 491
 QY 1111 RYSEDPVPLPSEF--DGYVAPLTCSPQPEYVQDPVRPQPSPREGLPAARPAQATLE 1168
 DB 492 RYSEDPVPLPSEF--DGYVAPLTCSPQPEYVQDPVRPQPSPREGLPAARPAQATLE 526
 QY 1169 RAKTLSPKNGVAVDF-----AFGAVENPEYL 1197
 DB 527 -----TAVVQNGQIYNNISLTAISKLPMDSRVQNSHSAVDNPEYL 566

RESULT 12
 Q8WYV0 PRELIMINARY; PRT; 412 AA.
 AC Q8WYV0;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical 44.7 kDa protein.
 GN P3659.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OK NCBI_TaxID=9606;
 RN [1]
 RA SEQUENCE FROM N.A.
 RA Huang Y., Zhou X.M., Zhang P.F., Jiang H.Q., Qin W.X., Zhao X.T.,
 RA Wan D.F., Gu J.R.;
 RT "Novel human cDNA clones with function of inhibiting cancer cell
 growth";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF318349; AAL55856.1; -
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR001245; Tyr_kinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00069; kinase.1.
 DR Pfam: PF02757; YLP.2.
 DR ProDom: PD000001; Euk_kinase.1.
 DR SMART: SM00219; TyrcKc.1.
 DR PROSITE: PS00108; EF_HAND; UNKNOWN.1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM.1.
 KW Hypothetical protein
 SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match 24.3%; Score 1697.5; DB 4; Length 412;
 Best Local Similarity 80.5%; Pred. No. 1.3e-120;
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4

QY 889 MALESILRRFTHQSDWVYGVYVWELMTFGAKPYDGIIPARELPDLEKGERLPQPPIC 948
 DB 1 MALESILRRFTHQSDWVYGVYVWELMTFGAKPYDGIIPARELPDLEKGERLPQPPIC 60
 QY 949 IDVYIMVYKCMWIDSECRPFRELVESEFRMARDPQRFVYIQNEIDGSPASPLDSTFYRSL 100
 DB 61 IDVYIMVYKCMWIDSECRPFRELVESEFRMARDPQRFVYIQNEIDGSPASPLDSTFYRSL 120
 QY 1009 LEDDDMDGLVDAAEYLYLPQGFPCPDPAAGAGMVAHRRSSSTRSGGDLTLGLEPSEE 106
 DB 121 LEDDDMDGLVDAAEYLYLPQGFPCPDPAAGAGMVAHRRSSSTRSGGDLTLGLEPSEE 180
 QY 1069 EAPSPAPSEAGSDVFDGDLGMAKGLQSLPTHPSPQLRYSEDPVPLPSETDGVY 112
 DB 181 EAPSPAPSEAGSDVFDGDLGMAKGLQSLPTHPSPQLRYSEDPVPLPSETDGVY 240
 QY 1129 APLTCSPOPEYVQDPVRPQPSPREGLPAARPAQATLERAKTISFGKNGVYKDVFAFG 1181
 DB 241 APLTCSPOPEYVQDPVRPQPSPREGLPAARPAQATLERAKTISFGKNGVYKDVFAFG 300
 QY 1189 GAVENPEYLYPQGGAPQ-----HPRA---FSFAFDNL 1221
 DB 301 GAVENPEYLYPQGGAPQ-----HPRA---FSFAFDNL 360
 QY 1221 YWVD-ODPPER-----GAPSPFKGTPTAEN 1245
 DB 361 YWVD-ODPPER-----GAPSPFKGTPTAEN 1410

RESULT 13
 Q64895 PRELIMINARY; PRT; 962 AA.
 AC Q64895;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Gag/v-erb-A,v-erb-B protein.
 GN GAG,v-erb-A,v-erb-B.
 OS Avian erythroblastosis virus.
 OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
 OC NCBI_Taxid=11661;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90206603; PubMed=1669616;
 RA Bruskini A., Jackson J., Bishop J.M., McCauley D.J., Schatzman R.C.;
 RT "Six amino acids from the retroviral gene gag greatly enhance the
 transforming potential of the oncogene v-erb-B";

RL Oncogene 5:15-24(1990).
 CC - SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC - SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 DR EMBL: X52209; CA36459.1; -
 DR EMBL: X52211; CA36459.1; JOINED.
 DR HSSP: P10828; 2NIL.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR000536; Hormone_rec_119.
 DR InterPro: IPR001723; Steroid_receptor.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR InterPro: IPR001628; Znf_C4steroid.
 DR Pfam: PF00104; hormone_rec; 1.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00105; Zf-C4; 1.
 DR PRINTS: PR00396; STROHOMONER.
 DR PRINTS: PR00047; STROIDPINGER.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR ProDom: PD000035; Znf_C4steroid; 1.
 DR SMART: SM00430; HOL1; 1.
 DR SMART: SM00399; Tyrc; 1.
 DR SMART: SM00399; Znf_C4; 1.
 DR PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
 KW Transcription regulation; Transferase; Tyrosine-protein kinase;
 KW Zinc finger.
 SQ SEQUENCE 962 AA; 109320 MW; 3C5AED791BEA95CE CRC64;

 Query Match 24.3%; Score 1653.5; DB 15; Length 962;
 Best Local Similarity 51.7%; Pred. No. 1e-116;
 Matches 358; Conservative 73; Mismatches 142; Indels 119; Gaps 18;

QY 1107 SPIGRSEDPYPLPSBTGTYAPLITCSQPEYVNOQPDVROPSPSPREGLPLAAPACAT 116
 DB 899 -----PVREDGFL-----PAPRYNQ--LMPKPKSTAVONQIYNYISLT 936

 QY 1167 -LERAKTSPGRKGVKQVFAFGCAVENPEYL 1197
 DB 937 AISKLPMSDRYQN-----SHTAVDNPPEYL 961

 RESULT 14
 Q85468 PRELIMINARY; PRT; 545 AA.
 AC Q85468;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Avian Erythroblastosis virus (TS34) v-erbB gene.
 OS Avian erythroblastosis virus.
 OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
 NC NCBI_TaxId=11861;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88217326; PubMed=2897102;
 RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
 RT "Common site of mutation in the erbB gene of avian erythroblastosis
 virus mutants that are temperature sensitive for transformation";
 RL Oncogene Res. 1:265-278(1987).
 DR EMBL: X06943; CA330024.1; -
 DR HSSP: P11362; IFGR.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF000069; pkinase; 1.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 545 AA; 60899 MW; 140CEB8CCA0F8A74 CRC64;

 Query Match 24.2%; Score 1645; DB 15; Length 545;
 Best Local Similarity 54.9%; Pred. No. 1.9e-116;
 Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

```

QY 396 PASPLDSTFYRLSLDDMDGLVDAEYLVPQGFCCDPAPAGAGVHRHRSSTRSG 1055
DB 414 LPSPTDSKFEYRTLMEEEMEDIVDADEVIVHQGF-----NSPST--- 454
QY 1056 GGDLTLGLFSESEAPRSPPL-----APSEGAGVDFDGLMGAKGLQSLPTDPSPLQ 1110
DB 455 -----SRTPLSLSLATSNNTXNTIDNGG-----H----- 481
QY 1111 RYSEDPTEVLPSETDGYVAPLTCSPQPEYVNPQDPVCPSPREGPPAPAPACATLER 1169
DB 482 -----PYREDGFL-----PAPYVNC--LMPKPKSTAVYQIYVYISLTAKSK 523
QY 1170 AKTLSPGKGVVKKDFAFGAGVENEPEYL 1197
DB 524 LPMDSRYQN-----SHSTAVDNEPEYL 544

RESULT 15
Q9WVF5 PRELIMINARY; PRT: 655 AA.
AC Q9WVF5;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., Wainhe N.J.;
RT "Alternative transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVITAC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J., Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D., Wainhe N.J.;
RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egfr transcripts encoding truncated receptor isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Akakawa T., Hara A., Fukumishi Y., Komoto H., Adachi J., Furukawa I., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada S., Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Asiburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H., Kuehl P., Lewis S., Matsumoto Y., Nikaide I., Pesole G., Quackenbush J., Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Wastio T., Sakai K., Oikido T., Furuno M., Aono H., Baladrelli R., Barish G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamaya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberees P., Nourjone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kontsuki S., Hayashizaki Y.;

```

```

RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL, AF124513; AAD44149.1; -.
DR EMBL, AF275365; AAG28047.1; -.
DR EMBL, AF275364; AAG28047.1; JOINED.
DR EMBL, AF275365; AAG28047.1; JOINED.
DR EMBL, AK004944; BAB33688.1; -.
DR EMBL, AK004893; BAB33641.1; -.
DR EMBL, AK004911; BAB33662.1; -.
DR MGI, MGI:95294; Egfr.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR002174; Furin-Like.
DR Pfam, PF01030; Recept_L_domain; 2.
DR SMART, SM00261; FU; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 21.8%; Score 1482.5; DB 11; Length 655;
Best Local Similarity 43.2%; Pred. No. 576-104;
Matches 276; Conservative 99; Mismatches 241; Indels 23; Gaps 7;

QY 11 LVALLPQAA--STQVCTGDMKLRLPASPETHLDMLRYGCGVYQGNLELYLPTN 68
DB 14 LITLCAAGALAEKRVQGTSTNRLTQIGTFEDHFLSLQRYNNCEVLLGNLEITYQFN 73
QY 69 ASLSFLQDIOEYQOYIKANSKFITGLQRLRIYRGTOLEFDNVALAVLNDGPNLNTTP 128
DB 74 YDLSLFKTIQEVAGVLIATLNTVERIFLENQITRGALYENTVALALSN----- 124
QY 129 VTGASPGGLRELOLSRLTEILKGVLIQRPOLCYQDITLWKDI----FKKNQALATLI 184
DB 125 -YGNRTGLRELPRNQLLEILIGAVRSNNPILCNMOTIQMDRIYQVAFNSMSMDL--- 180
QY 185 DYNRSRAHPGSPNCKSRCKGSESSDQSLRTVCGGCA-RKGLPLPDCCGEGCAAG 243
DB 181 -QSHPSCKPKDPCSPGNSCWGGEENCQKLTIKLCAQCSHRGRGSPSCCHNQCAG 239
QY 244 CGPHGSDCLAHFNHSGICELCPALVTYNTDTPFSMPPEGRYFGASCVACPYN 303
DB 240 CTGPRSDCLVQKQTDATKCTDCTPMLYNTPTTYQMDVNPESKYSFGATCYKCKRNY 299
QY 304 LSTDVSGSTLVCPHNOEVTADGTORCEKSKPCARVYGLGNEHLREYRAVTSANIOE 363
DB 300 VYTDHSGCVRAAGPDIYEV-EEIDIRCKCKDGGCRKXNGIGIGEFKDTLSINATYKH 358
QY 364 FAGCKKIRGSLAFIPESPDGPASNTAPLOEOLQVETLEITGYLYISAMPSPDLS 423
DB 359 FRYCTAIGDHLILVAKGDSFRTPLDRELEILKYAEITGFLLIQWPNWTDLH 418
QY 424 VFQNLQVIRGRILHNQAVSLTLOGISMLGRLSRLRELSGLALIHNTLCPVHTVPWD 483
DB 419 AFENLEIRGRTRKQHGQSLAVNGVINTSISGLRSLKEISDDVILISNRNLQVANTIMK 478
QY 484 QLFRRPHQALLHTANRPDECEVGEGLACHQICAGCHCKMGREPTQCVNCSQFLRQEEVEE 543
DB 479 KLFGRPNKTKIMNRAKDKAVNHNCPNLCSSSEGGWGPBPCVSCQVNSRKEVEK 528
QY 544 CAVTLOGLREYVNAHCLPCHPECPQNGSTYCGPAPADOCVACAHKDPFCVABCPSG 603
DB 539 CNILGGERFEVENSECIGCHPELCPANMITCTGRPDNDCIOCAHITDHPHCKTCPAG 598
QY 604 VKPDLSTYPIKFPDEBACOPCPINCTHSCVDLDYDGC 642
DB 599 IMGENNTL-VVKYADANNVCHLCANCTYGCAGPLOGC 636

```

Search completed: July 22, 2003, 09:00:06
 Job time : 53.3575 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.5887 Seconds

(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-103-117-14

Perfect score: 6847

Sequence: 1 MBLALCRWGLLALLPPGA.....TFKGTPTAENPEYLGIDVPY 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 20604715 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL 21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_prodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6232	91.0	1259	6	018735 canis fam1
2	3134	45.8	1209	1	09GX70
3	3106	45.4	1210	11	09EP98
4	2729.5	39.9	1137	13	09MEF6
5	2729	39.9	1165	13	09YH40
6	2313	33.8	1328	13	09YH40 xiphophorus
7	2042.5	29.8	1433	5	P79754
8	1776.5	25.9	419	4	09B1H9
9	1739	25.4	367	11	09UK79
10	1720	25.1	729	15	08R2X1
11	1718	25.1	567	15	08E712
12	1697.5	24.8	412	4	08E714
13	1633.5	24.1	922	15	08WV0
14	1645	24.0	345	15	08A895
15	1497.5	21.9	655	11	08A468
16	1481.5	21.6	643	11	09WVF5
					09ERV6
					09ERV6 mus musculu

17	1286	18.8	1193	5	09Y1X8	09Y1X8 ephydaria f
18	1192.5	17.4	1368	5	023821	023821 caenorhabd
19	1159	16.9	1717	5	026566	026566 schistosoma
20	1125	16.4	527	13	090836	090836 gallus gal.
21	992.5	14.5	478	11	09ES50	09ES50 rattus nor
22	941.5	13.8	599	13	09PSH2	09PSH2 gallus gal.
23	906	13.0	165	4	014256	014256 homo sapien
24	887	12.2	176	11	0923V5	0923V5 rattus nor
25	806.5	11.8	346	13	P11776	P11776 xiphophoru
26	778	11.4	435	5	08S2W1	08S2W1 drosophila
27	754.5	11.0	311	13	099162	099162 xiphophoru
28	746.5	10.9	1362	13	09PV24	09PV24 xenopus lae
29	723	10.6	149	6	09B6G6	09B6G6 oryctolagus
30	721	10.5	331	4	09BUD7	09BUD7 homo sapien
31	716	10.5	1671	5	09N1V5	09N1V5 biomphalari
32	691.5	10.1	1368	13	08UW85	08UW85 paraliichth
33	687	10.0	1418	13	093457	093457 scopthalmu
34	674.5	9.9	1369	13	08UW86	08UW86 paraliichth
35	667.5	9.7	1472	5	09U5A8	09U5A8 bombyx mori
36	661	9.7	1358	13	073798	073798 xenopus lae
37	643	9.4	1412	13	08UW84	08UW84 paraliichth
38	633.5	9.3	1418	13	08UW83	08UW83 paraliichth
39	632	9.2	1245	13	09YGH8	09YGH8 scopthalmu
40	629	9.2	1371	11	09QVW4	09QVW4 rattus sp.
41	614.5	9.0	2144	5	09VDY4	09VDY4 drosophila
42	608	8.9	987	11	091YW0	091YW0 mus musculu
43	606	8.9	935	4	096135	096135 homo sapien
44	605	8.8	987	11	099WR2	099WR2 mus musculu
45	587.5	8.6	1036	4	007912	007912 homo sapien

ALIGNMENTS

RESULT 1	018735	PRELIMINARY:	PRN: 1259 AA.
AC	018735		
DT	01-JAN-1998 (TREMBLrel. 05, Created)		
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Erdb-2.		
OS	Canis familiaris (dog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		
OX	NCBI_Taxid:9615;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Yokota H.		
RT	"cDNA cloning of erdb-2 from canine mammary gland."		
RU	Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AB008451; BAA23127.1; -		
DR	HSSP; P1362; 1EGK.		
DR	InterPro; IPR002048; EF-hand.		
DR	InterPro; IPR000494; EGFR_L domain.		
DR	InterPro; IPR000719; Euk_Pkinase.		
DR	InterPro; IPR002174; Furin-like.		
DR	InterPro; IPR001245; Tyr_Pkinase.		
DR	InterPro; IPR004019; YLP motif.		
DR	Pfam; PF00757; Furin-like; 1.		
DR	Pfam; PF00069; Pkinase; 1.		
DR	Pfam; PF01030; Recep_L domain; 2.		
DR	Pfam; PF02757; YLP; 2.		
DR	ProDom; PD000001; Euk_Pkinase; 1.		
DR	SMART; SM00261; FU; 3.		
DR	SMART; SM00219; TyKc; 1.		
DR	PROSITE; PS00018; EF_HAND; UNKNOWN 1.		
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.		
DR	PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.		
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.		
KW	ATP-binding; Transferase; Tyrosine-protein kinase.		
SO	SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;		

Query Match 91.0%; Score 6232; DB 6; Length 1259;
 Best Local Similarity 90.9%; Pred. No. 0;
 Matches 1153; Conservative 40; Mismatches 59; Indels 16; Gaps 4;

QY 1 MELALCRWGLLLALLPGMASTQVCTGDMKRLPASPEHLMDLRLHLYGQCVVQGNL 60
 DB 1 MELAAWCRWGLLLALLPGMASTQVCTGDMKRLPASPEHLMDLRLHLYGQCVVQGNL 60
 QY 61 ELTYLPNTASLFDLDIOEVGVYLIAHNOYRQVLOSLRVRTQOLFEDNYALAVLDNG 120
 DB 61 ELTYLPANASLFDLDIOEVGVYLIAHNOYRQVLOSLRVRTQOLFEDNYALAVLDNG 120
 QY 121 DPLNNFNFTVSFWLRVPEVSASHLE--QLRSLTEILKGVLLIORNQLCYQDTILMKDI 178
 DB 121 DPLEG-----GIPARCAAGGLREILQRLSLTEILKGVLLIORNQLCYQDTILMKDV 172
 QY 179 FHKNNQALTLIDTNRSPACHFCSMKGSKGSGESSDCSLRTVYAGGCAKGPPL 238
 DB 173 FHKNNQALTLIDTNRSPACHFCSMKGSKGSGESSDCSLRTVYAGGCAKGPPL 232
 QY 239 TDCHEQCAAGCTGPKHSDCLALHFNHSGICELHCPALVTYNTDFESMPNPEGRYTFG 298
 DB 233 TDCHEQCAAGCTGPKHSDCLALHFNHSGICELHCPALVTYNTDFESMPNPEGRYTFG 292
 QY 299 ASCVTACPIYNYLSTVVGCTLYVCLHNOEVTABDGTORCEKSPCARVCYGLAMEHLRE 358
 DB 293 ASCVTSCPIYNYLSTVVGCTLYVCLHNOEVTABDGTORCEKSPCARVCYGLAMEHLRE 352
 QY 359 VRAVTSANIQEFAGCKIFGSLAFIPESFDGDPASNTPLPDELOVPELLEETGLYI 418
 DB 353 VRAVTSANIQEFAGCKIFGSLAFIPESFDGDPASNTPLPDELOVPELLEETGLYI 412
 QY 419 SAMPDLPDLSVFONLQYIRGRILHNGAYSLTLQGLISWGLRLSRLSGALIHNT 478
 DB 413 SAMPDLPDLSVFONLQYIRGRILHNGAYSLTLQGLISWGLRLSRLSGALIHNT 472
 QY 479 HLCFVHTVPMDOFLFNPQOALLHTANREDECVBEGALACHQLCARGHCKMGPPIQCNCS 538
 DB 473 HLCFVHTVPMDOFLFNPQOALLHTANREDECVBEGALACHQLCARGHCKMGPPIQCNCS 531
 QY 539 QFLRQCEVEECRVLQGLPREVYNAHCLPCHPECOPONSVTCFGEADQCVACAHYKD 598
 DB 532 QFLRQCEVEECRVLQGLPREVYNAHCLPCHPECOPONSVTCFGEADQCVACAHYKD 591
 QY 599 PFCVACRCPGSKVPLSLMPIMKPFDEGACQCPINCTHSQVLDKCCPAEGRASPLT 658
 DB 592 PFCVACRCPGSKVPLSLMPIMKPFDEGACQCPINCTHSQVLDKCCPAEGRASPLT 651
 QY 659 SIVSAVGLLVVLGVVFGILIKRQCKIRKYMRRLLCETLVEPLTSGAMPNOAQM 719
 DB 652 SIVSAVGLLVVLGVVFGILIKRQCKIRKYMRRLLCETLVEPLTSGAMPNOAQM 711
 QY 719 RLKETEIRKVKVLSGAFGVYKCIAMPDEENKIPVAIKVRENTSPANKELIDEAV 778
 DB 712 RLKETEIRKVKVLSGAFGVYKCIAMPDEENKIPVAIKVRENTSPANKELIDEAV 771
 QY 779 VMAGVSPVSRLLGICLTSTVOLVTOAMPYGCCLLHVRENRGLSGODLLNCKMQLANG 838
 DB 772 VMAGVSPVSRLLGICLTSTVOLVTOAMPYGCCLLHVRENRGLSGODLLNCKMQLANG 831
 QY 839 MSYLEDEVLRDLAARNVLVKSNNHVKITDFGLARLLDIDETRYHADGGKVIKMALE 898
 DB 832 MSYLEDEVLRDLAARNVLVKSNNHVKITDFGLARLLDIDETRYHADGGKVIKMALE 891
 QY 899 SILRRFTTHQSDVMSYGTVMELTFGAKPYDGI PAEIDILEKGRLLQPPICITIDY 958
 DB 892 SILRRFTTHQSDVMSYGTVMELTFGAKPYDGI PAEIDILEKGRLLQPPICITIDY 951
 QY 959 MIMYKCMWIDSECPREFELVSESRMAPDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDD 1018
 DB 952 MIMYKCMWIDSECPREFELVSESRMAPDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDD 1011
 QY 1019 DMGDLVDAEELVQGGFFCPDPAFGAGGMVHHHRSSSTRSGGDLTLGLFSEBEAR 1078

DB 1012 DMGDLVDAEELVQGGFFCPDPAFGAGGMVHHHRSSSTRSGGDLTLGLFSEBEAR 1078
 QY 1079 SPLASEGAGSDVPFDGDLGMAAGKGLSLPHDSPLQRYSEDPVLPSETDGVAPLT 11:
 DB 1072 SPLASEGAGSDVPFDGDLGMAAGKGLSLPHDSPLQRYSEDPVLPSETDGVAPLT 11:
 QY 1133 CSPPEYNNQDPVRRQPPSPREGFLPAARPAATLER-----ATTLSGKNQVVKVDFAF 11:
 DB 1132 CSPPEYNNQDPVRRQPPSPREGFLPAARPAATLER-----ATTLSGKNQVVKVDFAF 11:
 QY 1194 GSAVENPEYLTPQGGAAQPPHPPAFSPAFNTLYYMODPBERGAPSTFGPTAENPE 12:
 DB 1192 GSAVENPEYLTPQGGAAQPPHPPAFSPAFNTLYYMODPBERGAPSTFGPTAENPE 12:
 QY 1254 YLGLDVPV 1261
 DB 1252 YLGLDVPV 1259

RESULT 2
 ID 09QX70 PRELIMINARY; PRT: 1209 AA.
 AC 09QX70;
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Epidermal growth factor receptor.
 GN EGFR.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RX MEDLINE=9025888; PubMed=2342466;
 RA Patch L.A.; Harris J., Raymond V.W., Blasband A.J., Lee D.C.,
 RA Earp H.S.;
 RT "A truncated, secreted form of the epidermal growth factor receptor is
 RT encoded by an alternatively spliced transcript in normal rat tissue";
 RL Mol. Cell. Biol. 10:12973-2982(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RA Patch L.A.;
 RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RA Gutteridge K., Dawson T.L., Earp H.S.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: M37394; A014008.1; -
 DR HSSP: P11362; IFCG.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_Pkinase.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; Kinase; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR PRINTS: PR00109; TYRKINASE.
 DR Prodom: PD000001; Euk_Pkinase; 1.
 DR SMART: SM00261; FU; 3.
 DR SMART: SM00219; TYRK; 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE: PS00109; PROTEIN KINASE DOW; 1.
 DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.
 KW ATP-binding; Receptor; Tyrosine-protein kinase.
 SQ SEQUENCE 1209 AA; 134891 MW; 96FEE7F6C1B773 CRC64;

Query Match 45.8%; Score 3134; DB 11; Length 1209;
 Best Local Similarity 49.7%; Pred. No. 7; 6e-229;

Matches 638; Conservative 166; Mismatches 359; Indels 120; Gaps 27;

```

QY 3 LAALCRWLLALLPFGA-ASTOYCTGDMKLRLPAPBETHLDMLRHLGGCGVQGNLE 61
Db 15 LAALCAAG-----GALKEKVCQGTSNRLTGLTFEDHFLSLQWENNCVAVLGNLE 66
QY 62 LTYLPTNASTLFLDIOEVQGYVLIANNOVQVEQLRLVGTQTLFEDNYALAVLDNG 121
Db 67 ITTVQRNYDLSFLKTIQEVAGYVLIANTVERIPLENLQIIRGNALYENTYALAVLSN-- 124
QY 122 PLNNFNNTVSFWLRVPKVSASHLEQLRSLTELLKGVLIQRNPOLCYOTITMKIIFHK 181
Db 125 -----VGTNKTGLREL-PRNNLOETILIGAVRSNNPILCNMTETQWRDIV-Q 169
QY 182 NNOALALLIDITNRS-RACHPCSPMCKSGRCWSESSEDCQSLTFETVCAGCA-RCKGELPT 239
Db 170 DVFLSNMSMDYQRHLTGCPKCDPCPNMGSCMGREBNQCQTLTKIICACQCSRRCRGRSBS 229
QY 240 DCHEQCAAGCTGPKHSDCLALCFHNSGICELHCPALVTYNTDTFESMNPGRYTFGA 299
Db 230 DCHNQCAAGCTGPRSDCLVCHRFDEATCKDCPELMYNPTTYQMDVNPBGKYSFGA 289
QY 300 SCVTACPNYLTSTVGSCTIYCPILHNOEVTAEADTORCEKSCPRCARVCYGLMEHLREY 359
Db 230 TCVKKCPNYYVTHGSCVRAKCGPDYEV-EDGVSKCKKCDGPKRYCNGIGIGEFKDT 348
QY 360 RAVTSANTQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEOLQVETLEITGYLYIS 419
Db 349 LSINATNIKHKRYCTAISGDLHIIPVAFKGDSPFRTPPLPRELEILIKTKETIGFLIIQ 408
QY 420 AMPDSLPLSVFQNLQVIRGRILHNGAYSLTLQGLGSMGLRSLLESGALIHNNH 479
Db 409 AMPKMTDLHAFEDLEIRKTKHGGPFLAVGLNITSIGLSLSEISDGVYISGNH 468
QY 480 LCFVHTVPMQDLFRNPQALHTANRPEDGCVQGLACHOLCARHGWGPGPTQVCNCG 539
Db 469 LCYANTIKWKKLFTGPNQKTKIMNRAEKOKATNNHCNPLCSSEGWGPEDVCSQCN 528
QY 540 FLRQGECEYGRVIGLPRVYANRHCLPCHPEQOPQNSVTFGPEADQCAVCAHYKP 599
Db 529 VSRGECVCKNILEGEPREVENSECIOCHPELPTQNMITGTGGPNCIKCAHYVGG 588
QY 600 PFCVARCPSGVKPLDSYMPIKFPDEEGACQPCPINTHSCVDLDKGGCAEGRASAF-LT 658
Db 589 PHCVKTCPSGIMGNNTL-VKRFADANNVCHLCHANCTYCAAPGKGC--QDPEPKLP 645
QY 659 STSAAVVGILLVVLGVVFGI-LIKRQOKIRRYTKRRLIQETELVEPLTPSCAMPNQAQ 717
Db 646 SIATGIVGGLFIVV-VALGIGLFMRRLQVRRKTLRLLQERELVEPLTPSGEAPNQA 704
QY 718 MRILKETELRKVKVLGSGAFGYKGIWIPDGENVKIPVAIKYLRNTPKANKETILDEA 777
Db 705 LRIIKETEFKIKIKVLGSGAFGYKGIWIPDGENVKIPVAIKYLRNTPKANKETILDEA 764
QY 778 YVAGVGSPPVSRLLGLCLTSTVQLVTLQMPYGCCLDHVAKENRGSGDGLNMCWQIAK 837
Db 765 YVAVSVNPHVCRLGLCLTSTVQLTQMLPYGCCLDYVREHKDNGISQYLINMCQIAK 824
QY 838 GMSYLEDVRLVHRDLAARNLVKSPNNHYKTTDPSGLARLLDIDTEYHAGGKPIKMAAL 897
Db 825 GSNYLEDRRLVHRDLAARNLVKTPQHVKTIDGLAKLGAEEKEVHAAGGVKPIKMAAL 884
QY 898 ESILRRRTHQSDVMSYGVTVLWELMTFGAKPYDGIIPAREIPDLLEKGEELPQPICTIDV 957
Db 885 ESILHRIYTHQSDVMSYGVTVLWELMTFGSKPYDGIIPASEISLLEKGEELPQPICTIDV 944
QY 958 YMIWVKCMMTIDSECRPRFELVSEFSHMADPRRFVYIQ-NEULGASLSDSTFRSLIE 1016
Db 945 YMIWVKCMMTIDSECRPRFELVSEFSHMADPRRFVYLVIGQDERMHLPSPTDNPFRALME 1004
QY 1017 DDDWGLDVAEEYLVPOGFFCPDPAFGAGMWHHHRSSSTRSGGDDTLGLESEBEA 1076
Db 1005 EEDWEDVVDADDEVILPQGGFF-----NEPST----- 1030

```

```

QY 1077 PRSPLAPEGASDPVFDGDLGMGAAXGLQSLPTHDPSPILQRYSEDPVLPSET--DGYV 113
Db 1031 SRTPLSLSLANSN-----SSTVACINNGSCRYKEDFLQRYSSDPVSTEDNIDTFL 106
QY 1135 APLTCSPOPEYVNPDPVPPQPPSPREGFLPAAPAGATLERAKTLSPGKNVAKVFAFG 115
Db 1087 -----PVPEYINQ-SVPKRPAGSVQNPVYHNGPLNH-----APGRLLHQN--PHS 112
QY 1195 GAVENPEYL-TPQGAAPQPHPPAPSPADNNTYMDQ-----DP-----PERG 123
Db 1130 NAVSNPEYLNPAQ-----PFLCSSGPDSSALWIKQSHQMSLDNPDYQODFFPKXA 116
QY 1238 APPSTFGKTPAENPEYLGIDVP 1260
Db 1181 KPNGIFKG-PIAENAEYLRVAPP 1202

```

```

RESULT 3
Q9EP98 PRELIMINARY; PRT; 1210 AA.
AC Q9EP98:
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor isoform 1.
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVETAC;
RA Reller J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Mathle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reller J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Schehl C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Mathle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF275366; AAG28045.1; -.
DR EMBL; AF275364; AAG28045.1; JOINED.
DR EMBL; AF275365; AAG28045.1; JOINED.
DR EMBL; AF275367; AAG24386.1; -.
DR HSSP; P11362; IFGK.
DR MG0; MG1; 95294; Egfr.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_Chr_Pkinase.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TYKc; 1.

```

DR PROSITE: PS00190; CYTOCHROME C, UNKNOWN_1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP_1.
 DR PROSITE: PS00111; PROTEIN_KINASE_DOM_1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR_1.
 KW ATP-binding; Receptor; Transferase.
 SQ SEQUENCE 1210 AA; 134840 MW; 62C0D021C9DE32E18 CRC64;

Query Match 45.4%; Score 3106; DB 11; Length 1210;
 Best Local Similarity 49.1%; Pred. No. 1e-226;
 Matches 629; Conservative 170; Mismatches 361; Indels 120; Gaps 24;

QY 11 LLLALPPGAA--STVCTGTDMLKLPASPETHLDMRPHLYOGGVQVQKLELYPTN 68
 DB 14 LLLALCAAGALEEKVCCGTSNRLTQLCTFEDHLSLQRYNCEVVLGALLETYYQRN 73
 QY 69 ASLSFLQDIQEVQGYVLINQVROVPIQRLIVRGTOIFEDYVALAVDNGDPLNNFN 128
 DB 74 YDLSFLKTIQEVAGYVLINLNTVERIPLENQIIRGNALYENTYALAILSNYG----- 126
 QY 129 FTVSFWLRVPKVSASHLEQ--RSLTEIKGGVLLQRNPOLCYODTLWKDI----FHKY 182
 DB 127 -----TNRGLRELPMRNQLEILIGAVRESNPNLCNMDDTIQMRDIQVNVMSN 175
 QY 183 NQIALTLIDNRSRACHPCSPMCKSRGWSESSDQSLRTVCAGGA-RCKGPLEPTDC 241
 DB 176 MSWDL----QSHPSSCPCKDPSCPNGSCWGGEENCQKLTIKIIQAQCSHRCRGRSPDC 231
 QY 242 CHEQCAAGCTGPKHSQCLALHFNHSGICEHLNPLVYNTDTEFSMPNREGRTFGASC 301
 DB 232 CHNQCAGCTGPRSDCLVQCKFQDEATCTKOTCPRLKYNPTTQMDVNPFGKXSFATC 291
 QY 302 VTACPYNYLSTDVGSCTVCPRLHNEVTAEDGTORCEKCSKPCARVCYGLMEHLREYRA 361
 DB 292 VKKCPNNYVVDHSGCVAPACPDYEV-EEDGIRCKCKCDGPRCKVCGIGISGFKDTLS 350
 QY 362 VTSANIQEFAGCKKIFGSLAFIPESFDDDPASNTAPLOPQLOVETLEETGLYISAM 421
 DB 351 IYATNIKFKKCTALISGDHLIPVAFKDSFTFRPDLRLEILIKYKELTGFLLOAM 410
 QY 422 PDSLPLSVFQNIQYIRGRILHNGAVSLTLOGLISWLGSLRELQSGLALIHNTHLG 481
 DB 411 PNMWTDLHAFELIIRGTRKHGQFSLAVGLNLTSLGLASLKEISGDVITIGNNLG 470
 QY 482 FVHTVPMQDLFNNPQALHTNREDECVGEGLACHOLCARHGWGPPTQVNCQSFL 541
 DB 471 YANTINMKKLFQTPQCKTKIMNNAEKCKAVNHVQNLCSSECGWGPBPDCQNVAS 530
 QY 542 RQCEVVEGCVLQGLPREYVNAHCLPCHPEGQFQNGSVTCFGEADQCVACAYXDPFF 601
 DB 531 RGRCEVEKCNILEGPREVEVSEICQHPCLPQAMNITCTGPGDNCTIOCAHYIDGPH 590
 QY 602 CVARCGSVKPDLSIMPTWKFPDEGACQCPINCTHSVDLDKGCFAEGRASPLTSIV 661
 DB 591 CVKTCPPAGIMGNNTL-VKXADANNVCHLCHANTYGCAPGQJGCEVWSPGPKISIA 649
 QY 662 SAVVGLLVVVLGVVFGI-LIKRQOKIRKXTRMRLQETELVEPLTPSGAMPQAMRI 720
 DB 650 TGIVGGLFIV-VALGIGLPMKRHHYRKATLRKLOERLVEPLTPSGARPQALRI 708
 QY 721 LKETELRKVKVIGSAGFTVYKGINIPGENVKIPVALKVLRENTSPRANKELIDEAYVM 780
 DB 709 LKETEFKKIKVLGSGAFGVYKGLMIPEGEVKYIPVALKEIREATSPRANKELIDEAYVM 768
 QY 781 AGVSPVYSRLIGLCTSTVOLVQIMVYGLLDHVENRRLSSODLLMWQOIAAGMS 840
 DB 769 ASVDNPHVCRLIGLITSTVOLITDLMVYGLLDYVRHKONISQYLLMWQVIAAGMN 828
 QY 841 YLEDVRLVHRDIAARNVLYKSPNHVKITDFGLARLLDIDEYHADGKYPKIMMALESI 900
 DB 829 YLEDRLVHRDLAARNVAVKTPQHVKITDFGLAKLLGAEXEYHAEGKYPKIMMALESI 888
 QY 901 LRRFTHQSQVWSYGVATWELMTFGAKYDGIIPARELPDLLEKERRPOPPICITDIYMI 960
 DB 901 LRRFTHQSQVWSYGVATWELMTFGAKYDGIIPARELPDLLEKERRPOPPICITDIYMI 960

DB 889 LRRITQSDVWSYGVATWELMTFGSKPYDGIIPASDISSILEKGERLPPOPICTIDYMI 941
 QY 961 MYKCMNIDSECRPRERELVSFSSMAARDPQRFVYIO-NEDLGPASPLDSTFYRSLBEDD 10:
 DB 949 MYKCMNIDSDSPKREILILEFSKAAARDPQRFVYIQQDERHHLSPDSNYRRLMBEED 10:
 QY 1020 MGDVDAEYLYVPOQGFCDPPARAGAGMHHRRSSSTRSGGDLTLGLEPSEEARPS 10:
 DB 1009 MDVVDADAEYLYVPOQGF-----NSPST-----SRT 10:
 QY 1080 PLAPSEGASDVFCDDLMGAKGLQSLPTHDPBLORYSEDPVPLPSET--GGYVAPL 11:
 DB 1035 PLSSLSATSN---NSIVACINRNGSCRVEDAFLOKRYSSDPTGATVEDNIDAFI--- 10:
 QY 1138 TCSPOPEYVNOQDVAPQPPSPREGPLPAPAPAGATLERAKTLSFGKGVYKDVFAFGAV 11:
 DB 1086 ---PVPEYVNOQ-SVPRKPAQSVQNPVYHNQPLHP-----ARGRLHYQN--PHSNAV 11:
 QY 1198 ENPEYL-TPOGGAAPQHPHPAPFAFNPVLYWPO-----DP-----PERGAPP 12:
 DB 1134 GNPETLNTAQ-----PTCLSGFNSPALWIKQSGHMSLDNPYQDFFPKETKPN 11:
 QY 1241 STFGKPTAENPEYLYGLDVP 1260
 DB 1185 GIFKG-PTAENAEYLRVAP 1203
 RESULT 4
 ID Q9W6F6 PRELIMINARY; PRT; 1137 AA.
 AC Q9W6F6;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Receptor tyrosine kinase (fragment).
 GN ERBB4.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NCBI_TaxID=9031;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=HINDRAIN;
 RX MEDLINE=99263203; PubMed=10328884;
 RA Dixon M., Lumsden A.;
 RT "distribution of neuregulin-1 (nrg1) and erbB4 transcripts in
 embryonic chick hindbrain.";
 RL Mol. Cell. Neurosci. 13:237-258(1999).
 DR EMBL; AF121963; AAD31764.1; -.
 DR HSSP; P11362; 1FGK.
 DR interpro: IPR000494; EGFR_L domain.
 DR interpro: IPR000719; Euk_Dkinase.
 DR interpro: IPR002174; Purin-like.
 DR interpro: IPR001368; TNFR_c6.
 DR interpro: IPR001245; Ty_Dkinase.
 DR interpro: IPR004019; YLP_moc1f.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF01030; Recep_L_domain; 1.
 DR Pfam; PF02757; YLP_3.
 DR PRINTS; PR00109; TYRKINASE.
 DR PRODOM; PD000001; Euk_Dkinase; 1.
 DR SMART; SM00261; FUY; 3.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP_1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM_1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR_1.
 DR PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.
 KW kinase; tyrosine-protein kinase.
 FT NON_TER 1
 SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;

Query Match 39.9%; Score 2729.5; DB 13; Length 1137;
 Best Local Similarity 47.3%; Pred. No. 3,66-198;
 Matches 539; Conservative 169; Mismatches 354; Indels 77; Gaps 24;

167 LCYODTILMKDIFHKNNMLALLIDTNRSAKCHGSPKCKSRWGSSSDQSLRTTVC 226
 3 LCFADTTHMODIVRNPNWASNFTLVFTNSSSGCRGCHSKCTG-RWGMPTENHCKTLTKTVC 61

227 AGGC-ARCKGRLPTDCHEQCAAGCTGPKHSDCLACTLHPNHSIGICELCPALVTYNTDTF 285
 62 AQCDCGRGCVGVSDCCRECGSGSGKDTDCACNPNFDSGACVATCCQTPTVYNTTF 121

286 ESMNPBGRITFGACVTAACPYNITSDVSCITVCPVLANOEVTAEDGTORCEKSPCA 345
 122 QLEHNNAKYTYGACVCKCPHNFTV-VDSSCVACPSKMEV-EENGIMCKRCPCTDICP 179

346 RVCYGLGMEHLREVAVTASANOEFAGCKIFGLSLFPSPDDPSNAPLQPELQV 405
 180 KACDSDIGTGLSVAQTVDSNIDKFINCTKINGNLILVTGHHDPYHTTAATPEKNT 239

406 FETLEITGYLISAMPDLSPLDSVPONLQYIRGRILHNGAVSLTGLGISMLGRSLR 465
 240 FQTVREITGYLNIOSWPEMNTDFRVFSNLVTIGRALYSGHLLILKQGITSLQPSLK 239

466 ELGSLALIHNTHLCPVHTVPMQDLFRNPQALHTANRPEDQVCEGLAQCLARGH 525
 300 QISAGNIYTNSNLCYHIVNMTSLSTPSQKTVIRNNKAENCTADGACNELSSDG 359

526 CWGPGPTQCVNCSQFLRGQECVECECRYLQGL-FREYVNAHRLCPHEPCOP-ONGSVTCFG 584
 360 CWGPGPDQCLCKKPIRGRTCTESCNLYDGEFRFANGSVCMEDDPCCKEDNMTCTCG 419

585 PEADQVACAHYKXPPPCVACSGGVAPDLSYMIKFPDEEGACQPCPINCTHSCVDDLD 644
 420 PGPHTCTKCFHFKQGPCVCEKCPDGLQGANSP-IFKVADEDECHCPHENCQGRGA 477

645 DKGC-----PABORASPLTSIVSAW-GILVVLVGVVFGILIRROQKIRKXT 692
 478 SHDCIYVPMTRQSTLPOHAR-TLP--IAAGVIGLFTIIVMGLTFVAVYRKSIR-KKA 533

693 MRLLQETELVEPLTPSGAMPNOQMILKETELRYKVYGSAGFGTYVKGWIPIPDENV 752
 534 LRRL-ETELVEPLTPSGTAPNOQLRIKETELRYKVYGSAGFGTYVKGWIPIPDENV 592

753 KIPVAIKVIRENTSPKANKELDAAVYMAVGVSYSRLIGLITSTVOLVQMLMYGL 812
 593 KIPVAIKILNETTPKXANVEPMDALIMASMDHPLVRLGVCLSPFIQVLVQMLPHGL 652

813 LDHYRENRGLSGODLLNMCQIAKMSYLEDVRLVHRDLAARNVLKSPNHKRTDFGL 872
 653 LDVYHEKDNIGSOLLNMCVQIAKGMVLEBRRLVHRDLAARNVLKSPNHKRTDFGL 712

873 ARLLDIETEHAGGVPIKMALESILRRFTHQSDVSVYTYWELMTFGAKPYDGI 932
 713 ARLLGDEKEYNAGGMPIKMALECIHAKRFTHQSDVSVYTYWELMTFGAKPYDGI 772

933 PAREIPDLEKGERLPOPICITIDVYMIWKCMMIIOSECPREIREIVSFESRRAPDQF 992
 773 PTREIPDLEKGERLPOPICITIDVYMIWKCMMIIOSECPREIREIVSFESRRAPDQF 832

993 VTIQNEP-LGPASLDSTFYSLLDEDDMGDLVDAEYLVPOQGFPCPDPAAGAGVNH 1051
 833 LVIOGDDRMKLPSFNSDKFQNLDEDEDLDDMDAEYLV-CAFNIPIPIYTSRTIDS 891

1052 RHSSSTRSGGDLTLGLEPSEEEAPRS-PLAP-SEGSGSDVFDODJGMAKGLQSLP 1108
 892 NRNGFYVRDGGVAAEGCV-PMYRABGCIIFEPAPVQAGATAELFEDTCNGTLRKQVATL 950

1109 THDPSPLQRYSEDPYPLBS-----ETDGYVAPLTCSPQPEYVNPDPVRQPPSPREG 1161
 951 AKEDSSTQRYSAPTVFIPIRVIRGELDEBDGYMTPRDKPKTIDYINPVENFVSRKNG 1010

1162 PLPAA-RPAGATLERAKTISFGKGVVQVF-----AFGGAIVENPEYLTPOGAAPQ 1212

DB 1011 DLQADNDEYHN-----ANQCPKXADEVNPBLYLNTFANTLENAEYL-----K 105

QY 1213 PAPPSPAPADNLYWDQDPPEGA--PSTFKGPT-----AENPEYL 1255

DB 1056 NULPEKAKKARDNPDYWNHSLPPRSTLQHPDYQEVSTKYFYKQNGRIRPIVAENPEYL 1114

RESULT 5

QYH40 PRELIMINARY; PRT; 1165 AA.

AC QYH40; 01-MAY-1999 (Tremblrel. 10, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE Receptor tyrosine kinase proto-oncogene.

GN XMRK.

OS Xiphophorus xiphidium.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;

OC Cyprinodontiformes; Poeciliidae; Xiphophorus.

OX NCBI_Taxid=8086;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=RIO PURIFICATION;

RX MEDLINE=98241172; PubMed=9582016;

RA Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,

RT "Activation of the Xmrk proto-oncogene of Xiphophorus by

RT overexpression and mutational alterations.";

RL Oncogene 16:1681-1690(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=RIO PURIFICATION;

RA Scharl M.;

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

EMBL; U53471; AAD1050.2; .

DR HSSP; P11362; IFGK.

DR InterPro; IPR000345; CYC_heme_bind.

DR InterPro; IPR000494; EGFR_L_domain.

DR InterPro; IPR000719; Euk_Dkinase.

DR InterPro; IPR002174; Furin-like.

DR InterPro; IPR001899; Gram_pos_anchor.

DR InterPro; IPR001245; Tyr_kinase.

DR Pfam; PF00757; Furin-like; 1.

DR Pfam; PF00069; Kinase; 1.

DR Pfam; PF01030; Recept_L_domain; 2.

DR PRINTS; PRC0109; TYRKINASE.

DR PRODOM; PD000001; Euk_pkinase; 1.

DR SMART; SM00261; FU; 3.

DR SMART; SM00219; TYRC; 1.

DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.

DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 2.

DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

DR KX ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.

SQ SEQUENCE 1165 AA; 129614 MW; 7E7EB3B8771A7AE CRC64;

Query Match 39.9%; Score 2729; DB 13; Length 1165;
 Best Local Similarity 45.5%; Pred. No. 4,1e-198;
 Matches 583; Conservative 163; Mismatches 382; Indels 154; Gaps 30

QY 1 MELALCRMGILLALLPPG-AAST-----OVCTGTDMKRLPASPEYTHDMLNHYGCGV 55

DB 4 LELLE-----LILLILISIGRCSTIDBRKVCQGTNSQNM---LDNHLYKMKXVYSGCNV 56

QY 56 VQGNLELTYPTNASSLFIQDQEVGVYLIHANOYRQVPLQRLAVGTOLFEDNYALA 115

DB 57 VLENLEITTYQEQDSLQSIQIEVGVYLIHANEVSTIPLVNLRLIGQNLIEGNTFL 116

QY 116 VLDN--GDLNNFNNTVTSFWLRVPKVASHLQQR--SLTEILKGVLIQNPQLCVQD 171


```

Db      117 WSNVQKNSS-----PDYVQVGLKQQLSNLLEILSGVKVSHNPLCNVE 163
Qy      172 TLKMDIFKXNQLALLIDNRSRACHPCSPMCKGRCSWSESSDCQLTRTVCAAGC- 230
Db      164 TINMWDIVDKTNSPTMNLIPHAFERQCKDPCGVNCSMAFGPHGCKFKTLCAECN 223
Qy      231 ARCKGFLPTDCCHECCAGCTGPKHSDCLACLFHNSGICLHRLPALVTYNTDFESMPN 290
Db      224 RRCKGKPIDCCNEHAGGCTGPRATDCLACDFDNDCTKDTCPPEKXIDIVSHQVVDN 283
Qy      291 PEGRYTFGASCVTACPYNYLSTDVGSCTLVPCPLHNOETAEADGTQREKSKPCARCYG 350
Db      284 PNKXTFGAACVKEGCSNVVTE-GACVRSACAGLEVD-ENGKRSCKPCDGVCPKXCDG 341
Qy      351 LGMHREVRATSNICQFAGCKTIFSGSLAFLESPFGDPASTAPLOPEQLOVFETLE 410
Db      342 IGISLSTIANVSNIGFSNCTINDIILNRSFEGDHYKIGPMDEHMLNLTIVK 401
Qy      411 EITGYLISAMPDLPDLVFQNLQVIRGRILHNGVSLTLQGLGISMGLRSLRELG 469
Db      402 EITGYLIVIMWPEMNTSLSFQNLLEIGRTTFSGSFVVVQVSHQWLGSLKVEVA 461
Qy      470 GLALIHNTLCCFHTVPMDLFNNPHQALHTANRPEDECVEGGLACHOLCARGHWP 529
Db      462 GNVILKNTPOLRYASTIMWRLLFRSEDSIEYDART-----ENQTCNNECSHDCWGP 514
Qy      530 GPTGCVNCGQFLRGEQVEEGRVLOGLEPREVYNAHCLPCHPEQPCQNGSVTCGPEADQ 589
Db      515 GPTMCVSLHVDGRGCVASCNLLOGEREQVDRGRCQCHQBELVTDTSTCGGEPAN 574
Qy      590 CVACAHKDPPEFCVACRCPGVKPLSYNPIKFPDEBACQPCPINCTHSCVDLDDKCP 649
Db      575 CSKCAHFQDGPQCIIPRCHGMLGDGDL-IKTVADKMQCQPCQNTQCSGSGR 633
Qy      650 AGRASPULSTVSAYVGLLVVGVGGLIKRQCKIRKTYTRRLLOETELVEPTPS 709
Db      634 GD-IVSHSLAVGLVGLITVYALVLVLRRLRK-RRTIRRLLOKEKEVEPTPS 691
Qy      710 GAMPQACMRILKETELRKVKVLGSGAFVYKGIWIPDGENYKIPAIKVLRENTSPXA 769
Db      692 GQAPQAFRLILKEIEFKKDRVLGSGAGTYKKLAMPDENIRIPAIKVLDEAISPXY 751
Qy      770 NKEILDEAYVMAGVSPVYSRLGICLTSTVQVLTQMPYGCCLIDHYENRGELSGQDL 829
Db      752 NOEVLDEAYVMAVDHPVCRLLGICLTSAVOLTOQLMPYGCCLIDYRQHGERICQWLL 811
Qy      830 NMCQIAGMSYLEDVRLVHRDLAARVLYKSPHVKITDQGLARLLDIDETRYHADGK 889
Db      812 NMCQIAGMNYLEBRHLVHRDLAARVLYKSPHVKITDQGLARLLDIDETRYHADGK 871
Qy      890 VPIKMALESILRRRFTHQSDVMSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGRLPQ 949
Db      872 VPIKMALESILQWYTHQSDVMSYGVTWELMTFGSKPYDGIIPAKETAVLNGERLPQ 931
Qy      950 PRICTIDYMMIMVCOMMIDSECRPRELVESEBRMADQRFVYVIONEDLGASPLDST 1009
Db      932 PRICTIEVYMIILKCMMDSPSRPRFRLVGEFQMARDSRYLVIG--NLPSPSBR 988
Qy      1010 FYRSLLEDMDGDLVDAEYLVPQCGFCPPAFAGAGMVAHHRSSSTRSGGDLTLGL 1069
Db      989 LFSRLSSD--DIVDADEYLL-----RYKRIN-RQGS----- 1018
Qy      1070 EPSEEARSPLASSEGAGSDVFGDGLGMAKAGLQSLPTHDSPLQRYSEDPTV-PLPS 1128
Db      1019 -----EPCTIPNGH-----PYRENSIALRYISDPTQNALER 1049
Qy      1129 ETDGVAPLTCSPQPEYVNCQDVRRPQ-----PSPRE-----GRLP-AARAGATLRR 1175
Db      1050 DLDG-----EYVNGPGETSSRLSDIYNPNVEDJTDGKGVYSISSQEAETNFSR 1099
Qy      1176 AKTISPCKGVAVKVFAGGAVENPEVLTTPQGAAPQHPPPAFSPAFDLVYVWDDPPE 1235

```

```

Db      1100 PEYLTNQNLSL----PLVSSSGMDPDY----QAG-----YQAAF-----LPQ 117
Qy      1236 RGAPSTFKGTPTAENPEYIGL 1257
Db      1135 TGALTNGMFLPAENLEYGL 1156

```

```

RESULT 6
P79754 ID P79754 PRELIMINARY; PRT: 1328 AA.
AC P79754;
DT 01-MAY-1997 (TRENBLREL 03, Created)
DT 01-MAY-1997 (TRENBLREL 03, last sequence update)
DT 01-JUN-2002 (TRENBLREL 21, last annotation update)
DE Etd33.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthopterygii; Acanthopterygii; Perciformes; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_Taxid=31033;
RN (1)
RP SEQUENCE FROM N. A.
RX MEDLINE=99177347; PubMed=10077531;
RA Gelliner K., Brenner S.;
RT "Analysis of 148 kb of genomic DNA around the wrt1 locus of Fugu
RT rubripes."; 9:251-258(1999).
RL Genome Res.
DR EMBL; AF056116; AAC34391.1; -.
DR HSP; P1362; 1FCX.
DR InterPro; IPR00494; EGFR_L domain.
DR InterPro; IPR000719; Euk_Dkinase.
DR InterPro; IPR002174; Tyrosine-like.
DR InterPro; IPR001245; Tyrosine-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR ProDom; PD000001; Euk_Dkinase; 1.
DR SMART; SM00261; FU; 3.
DR PROSITE; PS00107; Tyrosine_kinase; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR ATP-binding; TRANSFERASE.
SQ SEQUENCE 1328 AA; 148613 MW; A33039258B647E9 CRC64;

```

```

Query Match 33.8%; Score 2313; DB 13; Length 1328;
Best Local Similarity 40.48; Pred. No. 1,9e-166;
Matches 523; Conservative 154; Mismatches 417; Indels 200; Gaps 31

```

```

Qy      9 WGLLALILPP--GAASQ-----VCTGDMKLRPASPTHLDLRLHYQCCQVQGNLEL 62
Db      4 WRLTLCVASLRASSQTOEAVCPGQNGLSSTGSENGYNLNMKRYKCEIMANLEL 63
Qy      63 TYLPTNASLSLQDQGVGVYLIANQVRQVPLQRLRYRGQLFEDNYALVLDNGP 122
Db      64 TQISNNDPFLKTIREVTVYLIAMNHQEIPLQGLRVIRGNSLYERFALSV----- 117
Qy      123 LNNFNFTVSPWLKVPVYASHLQ--LRELTEILKGVYIQGNPQLQVQDTLMWKDIF 180
Db      118 -----FLNYPDQSPSLNQGLAMNLTILQGVQVILNKKYLRIGPVYVMDIT- 165
Qy      181 KNNQLATLIDNRSRACHPCSPMCKGRCSWSESSDCQLTRTVCAAGC-ARCKGRLPT 239
Db      166 RNDAPRIQFNGRGVCH--KSC-GNYCWGPGKQCOQLTITVCAPQCNDRCFTSPR 221
Qy      240 DCCHEGCAAGCTGPKHSDCLACLFHNSGICLHRCALTYNTDTFESPNPEGRTFCA 299
Db      222 DCHTECAAGCKGLDIDFCACLFLVDSGACVAPQCPQTLIYKQTFQMETNPAKYQYS 281
Qy      300 SCVTACPYNYLSTDVGSCTLVPCPLHNOETAEADGTQREKSKPCARCYGIGMHNRE 358
Db      282 ICVQGCPTHV-VDGSCVSVCPKPKMEV--ENGSGRQCELSGLCPKVCBGTGAE---Q 335

```

```

QY 359 VRAVTSANIQEPAGCKKIFGSLAPLPESEFDGDPASNTAPLOPELOQVETLEETIGLYLI 418
Db 336 ROTVSSNIDSFINTKIQGSLHFLVLTGDDPKVNPIDAKLEVFRTVREITDILNI 395
QY 419 SAWPDLPLDVFOVLQYIRGRILHNGVSLTLOGLSIMGLRLSELGSLALHNT 478
Db 396 QSWPELNDLSVFSLLTIGRSLFKRSLVMWRIPLTSLGRLSLREIDSGVYSQNA 455
QY 479 HLCFVHTVPMOOLFENPH-QALLHTANPEDECEGEGIACHQICARHCGMPGPTQVCNC 537
Db 456 HUCYHTVMTQOLFPGSRVNRANSLNSNRPMACVADRVDPLCSGSGCGPBDQCLSC 515
QY 538 SQFLAGQCEVECKVLOGLPREYVNAH-CLPCEPQOPONGSYTCGFPADQVCACAH 596
Db 516 RNSRHGTGVAGCHENSIPREFAGLNGCVACHCKPQGXKASCTGPGADCEMACTKX 575
QY 597 KDPFCVAPCPGVPPLDLYMPIMKFPDEGACQPCDINCHSCVDDDDGCAEGRAS 656
Db 576 RDGPYCMSSCPAGN-DGKGLIFKFRBCHCEPCQNTQCGSGGLNDC--LEAAR 631
QY 657 LITSVAVGILLVVLGVF-----GILIKRQCKIRKYTRRLLQETELVEPLTS 709
Db 632 LTISSGQITGIALGPAALIFCLVLFELGMLYHGLAIRRRAMRRLYESGESFEPLGP 690
QY 710 GAMPQACMRILKETELRKVKVULSGAFYVYKGIWIPDGENKIPAIKIVLENTSPKA 769
Db 691 GEKTKVHARILKSDLKIKPLSGVFGVSGKFWIPBETVYKIPALITDSSGRQT 750
QY 770 NKEILDEAVVAGVSPVSRLLICLTSTVQVLTQMLPYGCLLDHVENRGRGLSQDL 829
Db 751 FTEITDHLWSGSLDHPYIVRLGICPGTGLVLTQSSHGLEHTRCKTSLDQRL 810
QY 830 NMCQIAGMSYLDVRLVHRDLAARVULKSPHVKITPGLARLLIDETVHADGK 889
Db 811 NMCQIAGMSYLDVRLVHRDLAARVULKNDQVQVADLYPDCKKYVYSETK 870
QY 890 VPIKMALESILRRRTHQSDVMSYGVYVWELMTFAKPYDGIIPAREIPLEKGERLPQ 949
Db 871 TPIMMALESILFRYTHQSDVMSYGVYVWELMTFAKPYDGIIPAREIPLEKGERLPQ 930
QY 950 PPICTIDVYMTVMYKCMIDSECRPFRELVESESRMARDPQRFVIONEDLGPASPLDST 1009
Db 931 PAICTIDVYMTVMYKCMIDENIRPTFKELASDFRMAIDPRYLIVIRMG 980
QY 1010 FYRSLDEDDMDGLVDABEYLVPPQGFPCDDPARAGCMWHHRSSSTRSGGDLTLC 1069
Db 981 -----EDSGMGEFL-----RRGSE-----GLLEADL 1002
QY 1070 EPSEEARPRSLPABSEGAGSDVPDGLGMC--AAKGLQSLPTHDSPLQ----- 1116
Db 1003 EDEEE-----GLDDRFPATPSLQPSPEWSTSPQINSYMMTQL 1041
QY 1117 RYSEDPYPLPSETDGVAVPLTGFQF-ETVYNQ-----PDVROPSPREG 1161
Db 1042 RYD-----FAVSGGHHGYLPMSPVDITRQLWYQSRSLSVRTLPPRSARRSSRA 1095
QY 1162 PL--PAARPAATLERAATLSPGKNGVAVKDFAFGAVENPEVLTPOGGAAPQHPHAF 1219
Db 1096 ELCEDGACAGIFVVR-----FGSRGN-----PQGG----- 1122
QY 1220 SPAPDNLVYMDQDPPEKAPPSFTKGTPTAENPE 1253
Db 1123 -----QQRKLSTASSPSSFKTMADEDE 1146

```

```

GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Preygotia; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
OC Anopheles.
OX NCBI_TaxId=7165;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=SLA.
RA Lycet G.U.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor."
RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AJ301655; CAC35008.1; -.
DR HSB; P11362; IFCG.
DR InterPro; IPR000345; Cytochrome bind.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR01245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; kinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_kinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TYR; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR KX Receptor.
FT NON TER.
SQ SEQUENCE 1433 AA; 159585 MW; E3D9D896724F07 CRC64;

Query Match 29.8%; Score 2042.5; DB 5; Length 1433;
Best Local Similarity 32.7%; Pred. No. 7.6e-146;
Matches 474; Conservative 197; Mismatches 394; Indels 385; Gaps 37

QY 26 CTGTMDMLRLPASETHLMDLRLHYGQCCVYVQCNLELTYPTNASLFIQDIOEYGVYL 85
Db 1 CTGNGKMSVPANNEYVKNLRRYNTCYVDNLELTWQNTITDNLFIQHREYGVYL 60
QY 86 IAHNVQVPLQRLRIYRGTLFEDNYALAVLDNGDPLNNFNFTVSFMLRVKVSASHL 145
Db 61 ISLYDLPQVILPRQIIRGRTTK-----LKMGEAYGLF-----VSFSHM 101
QY 146 E--QLRSLTEILKGVLIQNPOLCYQDITLAKDI-FHKNNQALTLIDTNSRACHPCS 202
Db 102 NTLLEPALRDILGGSVGFNNVYLCHWKSINMEIILAPQTSQYTFNFSSEPRVCPC 161
QY 203 PMCKGSRGWSSESSDCSLTRTYCAGGCA--RCKGFLPTDCCHEGCACTGPKSDCLA 260
Db 162 PSCGCVG-CWGEAGHANCORSKLNCSPQCSGRFGGXPECCCLFAGGCTGPTSDCLA 220
QY 261 CLHNSGICELHCPALVYNTIDFESMPNPEGRYTFGASCYTAPEYNYLSTDVSGCTLV 320
Db 221 CKNFYDDGVCKQCEPQIYNFNFMPEPDPKAYAGATCYRKCP-ETLLDNGACVYK 279
QY 321 CPLHNGEVLTEDEGTQCEKSKCAVCGLGMEHLEBRATVTSANIQEPAGCKKIFGSL 380
Db 280 CPKGRVFNSE-----CVPKGVCPKTCGEGIVH-----SDNIGYKQCTTIEGSL 326
QY 381 AFLPESEFDGDPASNT-----APLQPELOQVETLEETIGLYISAMPDPLSLVFO 432
Db 327 EILDQSFQGVYVTFNSFGPRYIKIDPRLLEFVSFKYKLTIGFINIQAHHRFTLLNFR 386
QY 433 NLOVIRGRILHNGAY-SLTLOGISIMGLRLSELGSLALHNTLCVHTVPMOOLF 491
Db 387 NLEVVGRQLKENLFAVSYIVKTSKSELKSKIRVNSGIVLENSDLCFVEDIDWSEI 446

```

RESULT 7
Q9BIH9 PRELIMINARY; PRT; 1433 AA.
AC Q9BIH9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).

```

QY 492 FNNPQALHTANREDECEVGEGLACHQLCARHGWCPPTQVNCQPLRGQCEVECR 551
DB 447 KSSDHEWVWOKNRATCHEGEGECSCQCKAGCWGPPQCLCEKNVKKKGLDSC 506
QY 552 VLGGLPREY-VNARHCLPCHPEQCFQNGSVTCFGEADQCAVACHYDPPCVARCP 607
DB 507 ---SLPRLYSVDSKTCGDHCECKD---FCYGNEDNCSCNNVADGRFCVACETTK 558
QY 608 ---SGVXPPLSLMPTMKFPD----- 624
DB 558 HAMNGCTINCCKTCVCGCRGPRDITAPDGCISCDAIIGSDAKIERCLMKDSCDGYSD 618
QY 625 ---EEG----- 627
DB 619 VYLOEGEPLKOLSGKAVGRCKHCKCTGYFHEQFCQCTGYKKGEGECDEQDFYA 678
QY 628 ---ACQCPINCT---HSCVDL---D-----KCGPAEO----- 652
DB 679 NEETRICLPCHOEGCGHGLDHDHCECNLKLFGDDPYDNATTFTCVSNCPASHPYKRP 738
QY 653 ---RASPLTSYSAVAVGILLVVLGVFEGI---LTKRQOKIRK 690
DB 739 QEAGKIGPYCASDSMOSGLRIEPCQVXIVGSMVALLLCVVFGLAFVLFSSKAKKDA 798
QY 691 YTMRLLOETELVEPLTPSGAMPNOAQMRILKETELRKVYVLSGAFGYKGIWPDGE 750
DB 799 VKMTALAGCEDESEPLRPSNVGPNLTKRIKEAIRGGVLTGAGARGVFKGWMEGE 858
QY 751 NVKIPVAIKVLRENTSPRANKEILDEAVVAGVSPVYSRLGLCLSTQVLTQMLPYG 810
DB 859 SVKIPVAIKVLMWSSSSKEFELEAIYMASVEHPMLKILAVCMISOMMLITQMLPG 918
QY 811 CLLDHVRENGRLSGODLLNMCQIAKMSYLEVRLVHRDLAARNLVKSPNWKITDF 870
DB 919 CLLDVYRNKKKIGSKALLNWSQIARGMAYLEBRRLHRLAARNLVQTPSCVKITVF 978
QY 871 GLARLLDIDETEYHADGKVPILKMMALLESILRRFTHOSDVSAGVYVLMFEGAKPY 930
DB 979 GLAKLLDSDSEYRAAGKMPILKMTALECIRHRYFTSDVMAGITIMELLYGAPYE 1038
QY 931 GIPAREIPDLLEKGERLPORPITIDVYMIWVKWMTIDSECRPFRELVSEFSMARDPQ 990
DB 1039 NVPAKDVPELIEIGHKLPPQDICSIDVYCLISCVLDADARPTFKQLAETFAEKARDP 1098
QY 991 RRVVIONEDLGASLDSTFYRSLLEDODGDLY----- 1024
DB 1099 RYLMV-----PGDKFMRLPSTYNODEKDLIRTLAPVAMAAAAAAGASNDVPS 1149
QY 1025 ---DAEYLVPOQGFPCDPAPGAGWVHRHRSSTRSGGDLTLGLEPSEEARPS-- 1079
DB 1150 TIAETDEYLOPKTRSIMLPGPSA-----VEBS-DEMKSLR 1185
QY 1080 ---PLAP---SEGAGSYVFGDGLCMGAAGLQSLFTHDSPLQRYSEDPVLPESTG 1132
DB 1186 YCKDPLKPDDETGGKEV-----GVGGIR-----LMLPLDEDD 1219
QY 1133 YVAPLTCSPOEYVNPQVPRPQPSREGPLPAARPAATLERAKTLPCKNVAVDVFA 1192
DB 1220 YLMP-TCOSQ---NQS-----TPG---YNDLIG 1240
QY 1193 FGGAVENPEY-----TPGGAAPQPHPPAFSPAFDNLYWDQDPERGAPSTF 1243
DB 1241 VPASVDNPEYLMGSTOAILAGLAQSMG--PHTP-----PPNTP 1277
QY 1244 KGTPTAENPE 1253
DB 1278 NGMPTHHQSQ 1267

```

RESULT 8
Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
Q9UK79;

```

DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.", Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).
RN 12
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF177761; AAD56009.2; -
DR InterPro: IPR000494; EGFR_L domain.
DR InterPro: IPR002174; Furin-like.
DR Pfam: PF00757; Furin-like.
DR Pfam: PF01030; Recep_L domain; 1.
DR SMART: SMO0261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEGCIEB347E2D030C CRC64;

Query Match 25.9%; Score 1776.5; DB 4; Length 419;
Best Local Similarity 84.4%; Pred. No. 1,9e-126;
Matches 341; Conservative 5; Mismatches 37; Indels 21; Gaps 4

QY 1 MELALCRWGLLALLPFGAASSTVCTGDMKRLPASPTHLDMLRHYQGCQVQGNL 60
DB 1 MELALCRWGLLALLPFGAASSTVCTGDMKRLPASPTHLDMLRHYQGCQVQGNL 60
QY 61 ELTYLPNASTSPFQDIOEVQGYVLVHANCROYVQLRILVIGTQLFEDNVALLVNG 120
DB 61 ELTYLPNASTSPFQDIOEVQGYVLVHANCROYVQLRILVIGTQLFEDNVALLVNG 120
QY 121 DPLNNFNFTSFLRLR/KVKSAS---HLEQLSLREILKGVLLIORNPOLCYQDITLWK 176
DB 121 DPLNN-----TPPVGASFGGIRREIQLSLREILKGVLLIORNPOLCYQDITLWK 170
QY 177 DIFKNNQALATLIDITNRSRACHPCSPMKGSRGWSGSSDCQSLRTYVACAGCARCKGP 236
DB 171 DIFKNNQALATLIDITNRSRACHPCSPMKGSRGWSGSSDCQSLRTYVACAGCARCKGP 230
QY 237 LPTCCCHQCAAGCTGGRKSDCLACLFHNSGICELHCPALVTYNTDTFSMNPREGRTT 296
DB 231 LPTCCCHQCAAGCTGGRKSDCLACLFHNSGICELHCPALVTYNTDTFSMNPREGRTT 290
QY 297 FGASCVTACFPYVLTSTVGSCTVCPVHNOVETAAEDGTORCEKSPCARVCGLMEMH 356
DB 291 FGASCVTACFPYVLTSTVGSCTVCPVHNOVETAAEDGTORCEKSPCARVCGLMEMH 345
QY 357 REVDAAVTSANIQEPAGCKITFGSLAFIPESFDDPAPSNTAPLOP 400
DB 346 PPRPAAVFVPLRMQPG--PAHPVLSFLRPSWDVLSAFYSPLAP 387

```

RESULT 9
Q9R2X1 PRELIMINARY; PRT; 367 AA.
AC Q9R2X1;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;

RN (1)
 SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC027080; AAH27080.1;
 KW Hypothetical Protein.
 SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match 25.4%; Score 1739; DB 11; Length 367;
 Best Local Similarity 88.0%; Pred. No. 1,1e-123;
 Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 895 MALESILRRFRTHQSDVSYGVTTWELMTFGAKPYDGIIPAEIPDLLEKGERLPQPICT 954
 DB 1 MALESILRRFRTHQSDVSYGVTTWELMTFGAKPYDGIIPAEIPDLLEKGERLPQPICT 60
 QY 955 IDVYIMVYKCMWIDSECRPRFRELSEFSRMAPDPORFVYQNEIDLSPASPLDSTFVSL 1014
 DB 61 IDVYIMVYKCMWIDSECRPRFRELSEFSRMAPDPORFVYQNEIDLSPASPLDSTFVSL 120
 QY 1015 LEDDMGDLVDAEYLVYQOGFCFDPAPGAGVHHHRSSSTRSGGDLTLGLEPSEE 1074
 DB 121 LEDDMGDLVDAEYLVYQOGFCFDPAPGAGVHHHRSSSTRSGGDLTLGLEPSEE 180
 QY 1075 EAPRSPLAPSEGAGSDVFDGDLGMAAGLQSLPTHPSPLOQRYSEDPYPLASETDGY 1134
 DB 181 EAPRSPLAPSEGAGSDVFDGDLGMAAGLQSLPTHPSPLOQRYSEDPYPLASETDGY 240
 QY 1135 APLCSQPPEYVNOQDVPAPQPPSPREGPLPAAPAGATLEERAKTISPGKGVYVDAFAG 1194
 DB 241 APLCSQPPEYVNOQDVPAPQPPSPREGPLPAAPAGATLEERAKTISPGKGVYVDAFAG 300
 QY 1195 GAVNPEYLTROGGAAPQPPHPAPSPAFDNLTYWDDPPRPGAPSTFGTPTAENPEY 1254
 DB 301 GAVNPEYLTROGGAAPQPPHPAPSPAFDNLTYWDDPPRPGAPSTFGTPTAENPEY 360
 QY 1255 LGLDVPV 1261
 DB 361 LGLDVPV 367

RESULT 10
 Q86712 PRELIMINARY; PRT; 729 AA.
 AC Q86712;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE POLYPROTEIN.
 GN POLYPROTEIN.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
 OX NCBI_TaxID=11950;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; Pubmed=8152791;
 RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
 RA Johnson A., Beug H.;
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities";
 RL Oncogene 9:1307-1320(1994).
 DR EMBL: S69372; AAC60725.1; -;
 DR HSSP; P03322; 1A6S.
 DR InterPro; IPR000719; Euk PKinase.
 DR InterPro; IPR004028; Retro M.
 DR InterPro; IPR001245; Tyr PKinase.
 DR Pfam; PF00069; Kinase; 1.
 DR Pfam; PF02813; Retro M; 1.
 DR ProDom; PD000001; Euk_Kinase; 1.
 DR SMART; SM00219; TyKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.

DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 729 AA; 80649 MW; 84D269314EF1D53 CRC64;

Query Match 25.1%; Score 1720; DB 15; Length 729;
 Best Local Similarity 54.8%; Pred. No. 8.5e-122;
 Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 1;

QY 575 PONGSVTFCFGEAQCVAHAKPPCCVACRCSGVPRDLSTYPMIMFPEDEGACQCPPI 634
 DB 141 PEETAPPTGP--DHCKCAHFIDGPHCVACRPGVGENDTL-VWYVADANAVCQICHP 197
 QY 635 NCTHSCVDLDKCGCAEORASPLTSISAVY-GILVVLGVVFGILIKRQOKIRKRYTM 693
 DB 198 NCTRCCKRPGEGCP---NSGKTSIAAGVAGGLCLVVGGLGLLRRR-HIVKRTL 253
 QY 694 RRLDQETLVLEPLPSGSMQAOQMRILKETELKRVYVLSGAFYVYKQIMVPGENVK 753
 DB 254 RRLDQETLVLEPLPSGAPQOARILKETEFKRVYVLSGAFYVYKQIMVPEGEKVK 313
 QY 754 IPVAIKVRENTSPKANKELIDEAYVYVAGVSPYVRLGLCTSTVQVLTOLMPYGCIL 813
 DB 314 IPVAIKELRENTSRANKELIDEAYVYVADNPRVGCILGLCTSTVQVLTOLMPYGCIL 373
 QY 814 DHVRENGRLCSQDLLWCMQIAKMSYLEDVRLVHRDLAARVLYKSPNRYKITDFGLA 873
 DB 374 DYIREHKDNISQYLLWMCVQIAKKNVYLEERLVRHDLAARVLYKTPQHVKITDFGLA 433
 QY 874 RLDDIDEYVYVADGQVPIKMALESLRRFRTHQSDVSYGVTTWELMTFGAKPYDGI 933
 DB 434 KLLGADSEYVYVADGQVPIKMALESLRRFRTHQSDVSYGVTTWELMTFGAKPYDGI 493
 QY 934 AREIPDLLEKGERLPQPICTIDVYIMVYKCMWIDSECRPRFRELSEFSRMAPDPORFV 993
 DB 494 ASEISSVLEKERRLPQPICTIDVYIMVYKCMWIDSECRPRFRELSEFSRMAPDPORFV 553
 QY 994 VTC-NEDIGPASPDLSTFFRSLLEDMDGDLVDAEYLVYQOGFCFDPAPGAGVYVHR 105
 DB 554 VIQDERKHLPSPDTSKFRYTLMEEDMEDIVDADETLVPHQCF 598
 QY 1053 HRSSSTRSGGDLTLGLEPSEEARSPFL-----APSEGASDVFDGDLGMAAGLQSL 110
 DB 599 -NSPST-----SRFTPLSSLSATSNNSATNCID-----RNGQGH 631
 QY 1108 PTHDPSPLQRYSEDPYPLPSET--DGYPALTCSPQPEYVNOQDVPAPQPPSPREGPLPA 116
 DB 632 PVREDSFVQRYSSDPTGNFLEESIDQFL-----PAPERYNQ--LMPKKPS----- 675
 QY 1166 ARPAGATLEERAKTISPGKGVYVDAF-----AFSGAVNPEYL 1203
 DB 676 -----TAVYQNOIYNNISLITAIKLPMDSRVQNSHSTAVDNPEYL 715

RESULT 11
 Q86714 PRELIMINARY; PRT; 567 AA.
 AC Q86714;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE V-ERBB protein (Fragment).
 GN V-ERBB.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
 OX NCBI_TaxID=11950;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; Pubmed=8152791;
 RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
 RA Johnson A., Beug H.;
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities";

```

RL  Oncogene 9:1307-1320(1994).
DR  EMBL: S69372; AAC0727.1; -.
DR  HSSP: P13362; IFGK.
DR  InterPro: IPR000719; Euk_pkinase.
DR  InterPro: IPR01245; Tyr_pkinase.
DR  Pfam: PF00069; pkinase; 1.
DR  PRINTS: PR00109; TYRKINASE.
DR  ProDom: PD000001; Euk_pkinase; 1.
DR  SMART: SM00219; TyrcK; 1.
DR  PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR  PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
DR  PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW  Tyrosine-protein kinase.
FT  NON TER 1
SQ  SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF25E1 CRC64;

Query Match      25.1%; Score 1718; DB 15; Length 567;
Best Local Similarity 55.4%; Pred. No. 8,3e-122;
Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY  584 GREADQVACAHYKDPFCVACRPSGVKPDLSYMPFKPDEEGACOPCPINCHSCVDL 643
DB  1 GP--DHCMKCAHFIDQPHCVKACPGVGLGENDTL-VMKYADANAVCQLCPNCTRGCKGP 57
QY  644 DDKGCPAQRASPLSTYSAVV-GILLVVLGVVFGILIKRQCKIRKTYMRLLQETEL 702
DB  58 GLEGCP---NGSKTBSIAGVVGGLCLVWVGLGILYLR--HIVRKRLRLQEREL 113
QY  703 VEFLTPSGAMPQACMRILKETELAKVNLGSGAGTGYKIMIPDGENVKIPIATVLR 762
DB  114 VEPLTPSGEAPQAHRIKETEFKVKVLSGAGTGYKGLMIPSEKVIPIAIELR 173
QY  763 ENTSPKANKELDEAYVAVGVSPVSRLLGICLTSTVOLVTLQMLPYGCLLDHVENRGR 822
DB  174 EATSPKANKELDEAYVAVGVSPVSRLLGICLTSTVOLVTLQMLPYGCLLDHVENRGR 223
QY  823 LGSODLNMCMQIACMGVLEEDVRLVHEDLAARVLYKSPHVKITDPGLARLLDIDETE 862
DB  234 LGSQYLLNMCMQIACMGVLEEDVRLVHEDLAARVLYKTPQHVKITDPGLAKLGADEKE 293
QY  883 YHADGKVPKIKMMALESILRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLE 942
DB  294 YHAEKGKVPKIKMMALESILHRTYHQSDVMSYGVTVWELMTFGSKPFDGIPASISVLE 353
QY  943 KGERLPQPICTIDVYMIMVCMYIDSECRPRFELVSESRMARDPQRFVYIQ-NEDLG 1001
DB  354 KGERLPQPICTIDVYMIMVCMYIDADSRKPRFELVSESRMARDPRLVYIQGDERMH 413
QY  1002 PASPLDSTFYRSLDDMDGDIADAEYLVPOGFPCDPDAPAGAGVWHHRHSSSTRSG 1061
DB  414 LPSPTDSKFRYRLMEEDMEDIVDADEVLPVHQGF-----NSPST--- 454
QY  1062 GGDITLGLSEPEEAEPRSP-----APSEAGSDVFDGDLGMGAQKGLSLPTHDSPLQ 1116
DB  455 -----SRTPLLSLSLATSNSNATCID-----RNGQGHPRPREDSPVQ 491
QY  1117 RYSEDPVFLPSET--DGTVAPLTCSPQPEVYVNPQVPRPSPRSREPLPAARPAQTLE 1174
DB  492 RYSSDPVFLPSESLDDGFL-----PAPEYVNO--LMPKPS----- 526
QY  1175 RAKTISPGKNGVXKVF-----AFGAVENPEYL 1203
DB  527 -----TAMVQNOIYNNISLTAISKLPMSRYQNSHSTAVDNPVYL 566

RESULT 12
Q8WYVO PRELIMINARY; PRT; 412 AA.
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical 44.7 kDa protein.

GN  PP3659.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxId=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
RT  Wan D.F., Gu J.R.;
RT  "Novel human cDNA clones with function of inhibiting cancer cell
RT  growth.";
RL  Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR  EMBL: AF318349; AA55886.1; -.
DR  InterPro: IPR002048; EF-hand.
DR  InterPro: IPR000719; Euk_pkinase.
DR  InterPro: IPR001245; Tyr_pkinase.
DR  InterPro: IPR004019; YLP_motif.
DR  Pfam: PF00069; pkinase; 1.
DR  ProDom: PD000001; Euk_pkinase; 1.
DR  SMART: SM00219; TyrcK; 1.
DR  PROSITE: PS00018; EF_HAND; UNKNOWN 1.
DR  PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
KW  Hypothetical protein.
SQ  SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match      24.8%; Score 1697.5; DB 4; Length 412;
Best Local Similarity 80.5%; Pred. No. 1.9e-120;
Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY  895 MALESILRRFRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPICT 95.
DB  1 MALESILRRFRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPICT 60
QY  955 IDVYIMVCMYIDSECRPRFELVSESRMARDPQRFVYIQNEDLGPASPLDSTFYRSL 10;
DB  61 IDVYIMVCMYIDSECRPRFELVSESRMARDPQRFVYIQNEDLGPASPLDSTFYRSL 120
QY  1015 LEDDMGDLVAEEYLVPOGFPCDPDAPAGAGVWHHRHSSSTRSGGDLTGLPSEEE 10'
DB  121 LEDDMGDLVAEEYLVPOGFPCDPDAPAGAGVWHHRHSSSTRSGGDLTGLPSEEE 186
QY  1075 EAPRSPLAPSEAGSDVFDGDLGMGAQKGLSLPTHDSPLQRYSEDPVFLPSETGGYV 113
DB  181 EAPRSPLAPSEAGSDVFDGDLGMGAQKGLSLPTHDSPLQRYSEDPVFLPSETGGYV 240
QY  1135 APLTCSPPQPEVYVNPQVPRPSPRSREGLPAARPAQTLEPAKTLSPKNGVXKDYFAFG 119
DB  241 APLTCSPPQPEVYVNPQVPRPSPRSREGLPAARPAQTLEPAKTLSPKNGVXKDYFAFG 300
QY  1195 GAVENPEYLTQGGAAPOP-----HPPPA---FSPAFDNL 122
DB  301 GAVENPEYLTQGGAAALSPILLPSAPQSTTSITGTRTHQSGGHLPAKPSGHLRQRTQST 360
QY  1227 YYWD-QDPPER-----CAPSTFGKPTPAEN 1251
DB  361 WWTQCEPEGEQVRRSPDVSSGSHGLTSAGIKRWGSPPTTSGTCAHY 410

RESULT 13
Q64895 PRELIMINARY; PRT; 962 AA.
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Gag-V-erb-A-V-erb-B protein.
GN  Gag-V-erb-A-V-erb-B.
OS  Avian erythroblastosis virus.
OC  Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX  NCBI_TaxId=11861;
RN  [1]
RP  SEQUENCE FROM N.A.

```

EX MEDLINE=9020603; PubMed=1969616;
RA Bruckin A., Jackson J., Bishop J.M., McCarty D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the
transforming potential of the oncogene v-erb-B";
RT Oncogene 5:15-24(1990).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; X52209; CA36459.1; -
DR EMBL; X52211; CA36459.1; JOINED.
DR HSSP; P10828; 2NL.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000536; Hormone_rec_1ig.
DR InterPro; IPR001723; Strchnm_receptot.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR001628; Znf_Csteroi.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00105; zfc-C4; 1.
DR PRINTS; PR00398; STRDHOMONER.
DR PRINTS; PR00447; STROIPINGER.
DR PRODOM; PD000001; Euk_pkinase; 1.
DR PRODOM; PD000035; Znf_Csteroi; 1.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00219; Tyrc; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR RECEPTOR; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR ATP-binding; DNA-binding; Nucleic acid; Receptor; 1.
KW Transcription regulation; Tyrosine-protein kinase;
KW Zinc-finger.
SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E859CE CRC64;

Query Match 24.1%; Score 1633.5; DB 15; Length 962;
Best Local Similarity 51.7%; Pred. No. 1,4e-116;
Matches 358; Conservative 73; Mismatches 142; Indels 119; Gaps 18;

QY 547 VEEGRVQLPRE-YVNAR-HCLP-----CHPEQC 574
DB 354 IEKQESYLAFEHYIYRKNIHPHFSKLMKVADLRMGAYHASFILMKVECPTELS 413
QY 575 PONGSVTCFGEADQVCAHYKDPFCVACPSGVKPDLSYMPKPFDEGACQPCPI 634
DB 414 PQE-----VGP--DHQKCAHFIDGPHCVACPVAGLNDTL-VKYNADANAQCQLCHP 465
QY 635 NCHTSCVLDLDDKGPRAQASPLTYSVANV-GILLVVLGVVFGILIRROOKIRKXTM 693
DB 466 NCTRGCKGPGLEGP--NGSKTPSIAAGVAGGLCLVVGIGIGLYLRR-HIVKRTL 521
QY 694 RRLQETELVEPLTPSGAMPNQAQMRILKETELRKVYLSGAGFGVYVKGIMIPGSENVK 753
DB 522 RRLQERELVEPLTPSEARNQAHLILKETEKVKVYLGAGFGVYVKGIMIPGSENVK 581
QY 754 IPAVKYLRNTPSPKANKETLDEAYVMAVAGSPVYSRLIGICLTSTVQLTOLMEYGCIL 813
DB 582 IPAVKELREATSPKANKETLDEAYVMAVAGSPVYSRLIGICLTSTVQLTOLMEYGCIL 641
QY 814 DHYRENRGRSGODLNMCMQIAKMSYLEDRVLYHRDLAARNVLYKSNHYKITDEGLA 873
DB 642 DYIREHKDNIGSYLLNMCVQIAKGMVLEBRNHRDLAARNVLYKTPQHYKITDFELA 701
QY 874 RLIDIDETEHADGKVPYIKMMALLESILRRRFTHOSDVMSYGVTVWELMTFGAKPYDIP 933
DB 702 KQGADEKEHYHAGGVPIKMMALLESILRHYTHOSDVMSYGVTVWELMTFGSKYDIP 761
QY 934 AREIPPLEKGEPLPPICITDVMYIMKCMIMISECPRRRELVSFPMARDPQGFV 993
DB 762 ASSTSVLEKGEPLPPICITDVMYIMKCMIMISECPRRRELVSFPMARDPQGFV 821
QY 994 VIO-NEDIGPASPGLSTFYRSLLEDMDGLVDAEYVLPQGFPCPDPAAGAGMVRH 1052
DB 822 VIOGDERMHLPSFTDSKPYRTIMEEDVEDIVDADEYLVPHGGF----- 866

QY 1053 HRSSSTRGGDILLGLPSEEAAPRSEAPLAPSEAGSDVFDGLGMAAGLQSLPTHP 111
DB 867 -NSPT-----SRTPLLSSLSATSN-----NSATCICIRNGCH-- 896
QY 1113 SPLQRYSEDPVPUPSETDGVAPRLTCSPOPEVYNQPDVPOPPSPREGPLPAAPAGAT 117
DB 899 -----PVEDGDL-----PAPEYVNO--LMPKAPSTAMONQIYNYISLT 936
QY 1173 -LEPAKTLSPKNGVNVKDVPAFGAVENPEYL 1203
DB 937 AISKLPMDSRYN-----SHSTAIVNPEYL 961

RESULT 14
085468
ID 085468 PRELIMINARY; PRT; 545 AA.
AC 085468;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Avian Erythroblastosis virus (T834) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=2897102;
RA "Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:265-278(1987).
DR EMBL; X06943; CA30024.1; -
DR HSSP; P1362; 1FGX.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR PRODOM; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR ATP-binding; Tyrosine-protein kinase.
KW Transcription regulation; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 60899 MW; 140DCB8CA0F8AF4 CRC64;

Query Match 24.0%; Score 1645; DB 15; Length 545;
Best Local Similarity 54.9%; Pred. No. 2,7e-116;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15

QY 584 GPEADQVCAHYKDPFCVACPSGVKPDLSYMPKPFDEGACQPCPICTHSCVDL 643
DB 1 GP--DHQKCAHFIDGPHCVACPVAGLNDTL-VKYNADANAQCQLCHPCTGCKGSP 57
QY 644 DDKGPAEGRASPPLTYSVANV-GILLVVLGVVFGILIRROOKIRKXTMRLLQETEL 702
DB 58 GLEGP--NGSKTPSIAAGVAGGLCLVVGIGIGLYLRR-HIVKRTLRLQEREL 113
QY 703 VEPITSGAMPNQAQMRILKETELRKVYLSGAGFGVYVKGIMIPDGENVKIPIVKYL 762
DB 114 VEPITSGAPNQAHLIRKETEFKVKVYLGAGFGVYVKGIMIPDGENVKIPIVKYL 173
QY 763 ENTPSKANKETLDEAYVMAVAGSPVYSRLIGICLTSTVQLTOLMEYGCILDHRENGR 822
DB 174 EATSPKANKETLDEAYVMAVAGSPVYSRLIGICLTSTVQLTOLMEYGCILDHRENGR 233
QY 823 LGSQDILNMCMQIAKMSYLEDRVLYHRDLAARNVLYKSNHYKITDEGLA 882
DB 234 IGSQVYLLNMCVQIAKGMVLEBRNHRDLAARNVLYKTPQHYKITDFELA 293
QY 883 YHADGKVPYIKMMALLESILRRRFTHOSDVMSYGVTVWELMTFGAKPYDIPAREIPDLE 942
DB 294 YHAEQKVPYIKMMALLESILRHYTHOSDVMSYGVTVWELMTFGSKPYDIPAREIPDLE 353

```

QY 943 KGERLPPOPICTIDVYIMVCMKMSIDSECRPREPLVSEFSRMRDQRFVIO-NEDLG 1001
DB 354 KGERLPPOPICTIDVYIMVCMKMSIDSECRPREPLVSEFSRMRDQRFVIO-NEDLG 413
QY 1002 PASPLDSTFFYRSLBEDDMGDLVDAEYLVQOGFCFCDPAAGAGVHHRRSSSTRSG 1061
DB 414 LPSPTDSKFYRTLMEEDMEDIVDAEYLVHQGF-----NSPST---454
QY 1062 GGDULTGLEPSESEFAPRSP-----APSEGAGSVDFDGLGMAKGLQSLPHIDPEPLQ 1116
DB 455 -----SKRLPSSLSATSNNSATNCIDRNGS-----H-----481
QY 1117 RYSEDPYPLPSEITDGYAPLPTCPOPEYVQNPVDRPOPSPREGPPAARPAAGT-LER 1175
DB 482 -----PYREDFGL-----PAPRYVQ--LMPKPESTAVQVQIYNYSLTATSK 523
QY 1176 AKTLSPGKGVKXVDFAFGCAVENEXL 1203
DB 524 LPMDSRYQN-----SHSTAVDNEYL 544

RESULT 15
OQWVF5 PRELIMINARY: PRT: 655 AA.
AC OQWVF5;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C.,
RA Lampard A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maibhe N.J.;
RT "Alternative transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors.";
RT Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTA;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearisall R.S., Green P.J., Yee D., Lampard A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maibhe N.J.;
RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egfr transcripts encoding truncated receptor isoforms.";
RT Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN 13
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RC MEDLINE=21085660; PubMed=11217851;
RA Kawai T., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pedole G., Quackenbush J.,
RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Kono H., Badarrelli R., Barsch G.,
RA Blake J., Boffelli D., Bojunga N., Carminci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guernicelli S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

```

```

RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RT Nature 409:685-690(2001).
DR EMBL: AF124513; AAd4149.1; -
DR EMBL: AF275366; AAG38047.1; -
DR EMBL: AF275364; AAG38047.1; JOINED.
DR EMBL: AF275365; AAG38047.1; JOINED.
DR EMBL: AK004944; BAB23688.1; -
DR EMBL: AK004883; BAB23641.1; -
DR EMBL: AK004911; BAB23652.1; -
DR MGD: MGI:95294; Egfr.
DR InterPro: IPR000494; EGFR_L domain.
DR InterPro: IPR002174; Furin-like.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recept_L domain; 2.
DR SMART: SMO0261; Fu; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

```

```

Query Match 21.9%; Score 1497.5; DB 11; Length 655;
Best Local Similarity 43.3%; Pred. No. 5,7e-105;
Matches 280; Conservativity 100; Mismatches 214; Indels 33; Gaps 8

```

```

QY 11 LLLALLPFGAA--STOYCTGDMKLRLPAPBETHLMDLRLYGGCGVQGNLETLPTN 68
DB 14 LLLALLPFGAA--STOYCTGDMKLRLPAPBETHLMDLRLYGGCGVQGNLETLPTN 73
QY 69 ASLSFLODIOGVGYVLIHANOVOVELORLRIYRGTLFEDNYALAVLDNGPILNNFN 128
DB 74 YDLSFLTKIQGVAGVYLIANTVERIPLENLQIRGNALVENTYALALISNYG----- 126
QY 129 FTVSFWLRVPVKSASHLEQL--KSLTEILKGVLIQANPOLCTQDTITLMCDT----FKKN 182
DB 127 -----TNRGTGRELEPMRLQELILGAARFSNNPILCMWDITQMDIVQNVFMSN 175
QY 183 NQLATLIDITNRSACHPCSPMCKGRSCWGESEDDCSLRTVCAGGCA-RCKGPLPTDC 241
DB 176 MSMDL-----QSPSSCKCKDPSCKNGSCWKGCGENCKLTKIICAGQCSKRCGRSPSC 231
QY 242 CHEQCAAGCTGPKKASDCLALHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASC 301
DB 232 CHNCAAGCTGPKKASDCLALHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASC 291
QY 302 VTACPVYVLTSDVSCCTLVCPRLHNOEYTAEDGTORCKSKPCARVCYGLMHLREYVA 361
DB 292 VKKCPRIYVYVTDHSCCYRACGPDYEV--EDGIRKCKCKDGPCKVCKNGIGIEFPDITLS 350
QY 362 VTSANIGEPAGCKKIFGSLAFLPSEFEDGDSANTAPLOPEQLQVFTLEITGYLYISAM 421
DB 351 INATNIGFKCTAISGDLHLPLVAFKGDPSFTPLDPRELEILTKVETITFLDIQMW 410
QY 422 PDSLPLDSVQNLQVIRGRILNHGASVLTGIGISMLGIRSLRELGSGLALHNHTHLC 481
DB 411 PDNWTDLHAFENLEITGRCKQGQSLAVGANTISLGRSLKESIDSDVYIISGRNLC 470
QY 482 FVATVPMDQLFRNHQALLHTANRPEDEGESEGLACHQICARHCWGPPPTCVNCSQPL 541
DB 471 YANTINKKLFGRPNQTKIMNRAREDKCAVHVNCPLSSSGCGGPPRCDVSCQNVNS 530
QY 542 RGQCGVECKVLQGLPREVYNAHCLPCBPCCOPQNGSVTCGPEZDQCVACAHYKDPDF 601
DB 531 RGRCEVCKVILGEPEPEFENSECICQHECCLPQAMNITITCGRGDNCICQAHYIDGPH 590
QY 602 CVARCPGVKPDLSYPIWKFPEEGACQPCPINCCHSCVTDLDDKCC 648
DB 591 CVTKCPAGINGENNTL-VKTYALANNVCHLCHANCYTGCGAGFGLQGC 636

```

Search completed: July 22, 2003, 09:00:23
Job time : 54.5887 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:47:04 ; Search time 10.642 Seconds

(without alignments)

4891.279 Million cell updates/sec

Title: SEQ4-695-709-12

Perfect score: 6815

Sequence: 1 MELALCRWGLLALPFGA.....TFKQTPNENPEYLGLDVEY 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt 40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6640	97.4	1255	1 ERB2_HUMAN	P04626 homo sapien
2	5849	85.8	1257	1 ERB2_FAT	P06494 rattus norv
3	5833.5	85.6	1254	1 ERB2_MESAU	Q60553 mesocricetu
4	3065	45.0	1210	1 EGFR_HUMAN	P00533 homo sapien
5	3039	44.6	1210	1 EGFR_MOUSE	Q01279 homo sapien
6	2887.5	42.4	1308	1 ERB4_HUMAN	Q15303 homo sapien
7	2868	42.1	1308	1 ERB4_RAT	Q62856 rattus norv
8	2614.5	38.4	1167	1 XMRK_XIPMA	P13388 xiphophorus
9	2341.5	34.4	1342	1 ERB3_HUMAN	P13388 xiphophorus
10	2269.5	33.3	1339	1 ERB3_RAT	P62799 rattus norv
11	1897	27.7	1426	1 EGFR_DROME	P04412 drosophila
12	1659.5	24.5	634	1 ERBB_ALY	P00534 avian leuko
13	1623	23.8	604	1 EGFR_AYER	P00534 avian leuko
14	1595	23.4	703	1 EGFR_CHICK	P13387 gallus eryth
15	1558	22.9	540	1 ERBB_AYEU	P13387 gallus eryth
16	1243	18.2	1332	1 LIT2_CABEL	P24348 caenorhabdi
17	1142.5	16.8	1245	1 ERB2_MOUSE	P70424 mus musculu
18	700	10.3	1363	1 ILPR_BRAIA	Q02466 brachyosco
19	666	9.8	1383	1 INSR_RAT	P15127 rattus norv
20	665.5	9.8	1372	1 INSR_MOUSE	P15208 mus musculu
21	663	9.7	1300	1 INSR_MOUSE	Q99414 mus musculu
22	662	9.7	1382	1 INSR_HUMAN	P06213 homo sapien
23	658	9.7	1477	1 HTRX_HYDAT	Q25197 hydra atten
24	657.5	9.6	1300	1 IRR_CAVPO	Q25197 hydra atten
25	655	9.6	1607	1 MIFR_LYMT	Q25410 lymanaea sta
26	654	9.6	1297	1 IRR_HUMAN	Q14616 cavia porce
27	602	8.8	1367	1 IGR_HUMAN	P08065 homo sapien
28	591	8.7	1390	1 INSR_AEDAE	Q93105 aedes aegypt
29	590	8.7	2146	1 INSR_DROME	P09208 drosophila
30	584	8.6	1373	1 IGR_MOUSE	Q60751 mus musculu
31	580.5	8.5	1370	1 IGR_RAT	P24062 rattus norv
32	572	8.4	987	1 EPB4_HUMAN	P54760 homo sapien
33	565	8.3	1114	1 RET_HUMAN	P07949 homo sapien

34	563.5	8.3	984	1 EPB1_CHICK	Q07494 gallus gall
35	562.5	8.3	977	1 EPB2_MOUSE	Q03145 mus musculu
36	560.5	8.2	984	1 EPB1_RAT	P09759 rattus norv
37	554.5	8.1	984	1 EPB1_HUMAN	P54762 homo sapien
38	552.5	8.1	976	1 EPB2_HUMAN	P29317 homo sapien
39	552	8.1	902	1 EPBB_XENLA	Q91736 xenopus lae
40	552	8.1	987	1 EPB4_MOUSE	P54761 mus musculu
41	544.5	8.0	985	1 EPBA_XENLA	Q91571 xenopus lae
42	544	8.0	757	1 HT16_HYDAT	P53356 hydra atten
43	544	8.0	1068	1 FAK1_XENLA	Q91738 xenopus lae
44	542.5	8.0	1053	1 FAK1_CHICK	Q00944 gallus gall
45	532.5	7.8	1052	1 FAK1_MOUSE	P34152 mus musculu

ALIGNMENTS

```

RESULT 1
ID ERB2_HUMAN STANDARD; PRT; 1255 AA.
AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE surface receptor HER2) (MDN 19).
OS ERB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=66118663; PubMed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA Saito T., Toyoshima K.;
RT "Similarity of protein encoded by the human c-erb-B-2 gene to
RT epidermal growth factor receptor.";
RL Nature 319:230-234(1986).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=66070181; PubMed=2999974;
RA Cossens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA McGrath J., Seeburg P.H., Liberman T.A., Schlessinger J.,
RA Francis U., Levinson A., Ullrich A.;
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene.";
RL Science 230:1132-1139(1985).
RN (3)
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=66016729; PubMed=2995967;
RA Semba K., Kanata N., Toyoshima K., Yamamoto T.;
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a
RT human salivary gland adenocarcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN (4)
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT "Characterization of a new allele of the human ERB2 gene by allele-
RT specific competition hybridization.";
RL Genomics 15:426-429(1993).
-1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
ALPHA AND AMPHIREGULIN.
-1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
-1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
(POTENTIAL).
-1- SUBCELLULAR LOCATION: Type I membrane protein.

```

CC -1- PFM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES (BY SIMILARITY)
 CC -1- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN
 CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
 CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
 CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M1767; AAA55808.1; -
 DR EMBL: M1761; AAA55808.1; JOINED.
 DR EMBL: M1762; AAA55808.1; JOINED.
 DR EMBL: M1763; AAA55808.1; JOINED.
 DR EMBL: M1764; AAA55808.1; JOINED.
 DR EMBL: M1765; AAA55808.1; JOINED.
 DR EMBL: M1766; AAA55808.1; JOINED.
 DR EMBL: M1730; AAA5493.1; -
 DR EMBL: M12036; AAA55978.1; -
 DR EMBL: X0363; CAA27060.1; -
 DR PIR: A25491; A25491.
 DR PIR: A24571; A24571.
 DR HSP: P1362; IFGK.
 DR Genew: HGNC:3430; ERBB2.
 DR MIM: 164870; -
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR Pfam: PF02757; YLP; 2.
 DR Pfam: PF000001; Euk_pkinase; 1.
 DR SMART; SM00261; Fy; 3.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP_1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR_1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM_1.
 DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transmembrane; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Polymorphism.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
 FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 653 675 POTENTIAL.
 FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 720 987 PROTEIN KINASE.
 FT NP_BIND 726 734 ATP (BY SIMILARITY).
 FT BINDING 753 753 ATP (BY SIMILARITY).
 FT ACT_SITE 845 845 BY SIMILARITY.
 FT DISULFID 195 204 BY SIMILARITY.
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 220 227 BY SIMILARITY.
 FT DISULFID 224 235 BY SIMILARITY.
 FT DISULFID 236 244 BY SIMILARITY.
 FT DISULFID 240 252 BY SIMILARITY.
 FT DISULFID 255 264 BY SIMILARITY.
 FT DISULFID 268 295 BY SIMILARITY.
 FT DISULFID 299 311 BY SIMILARITY.
 FT DISULFID 315 331 BY SIMILARITY.
 FT DISULFID 334 338 BY SIMILARITY.
 FT DISULFID 511 520 BY SIMILARITY.
 FT DISULFID 515 528 BY SIMILARITY.
 FT DISULFID 531 540 BY SIMILARITY.
 FT DISULFID 544 560 BY SIMILARITY.

FT	DISULFID	563	576	BY SIMILARITY.
FT	DISULFID	567	584	BY SIMILARITY.
FT	DISULFID	587	596	BY SIMILARITY.
FT	DISULFID	600	623	BY SIMILARITY.
FT	DISULFID	626	642	BY SIMILARITY.
FT	DISULFID	630	642	BY SIMILARITY.
FT	MOD_RES	1139	1139	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1248	1248	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	68	68	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	124	124	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	187	187	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	259	259	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	530	530	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	571	571	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	629	629	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT	654	654	I -> V.
FT	VARIANT	655	655	/FTID=VAR_004077.
FT	VARIANT	655	655	I -> V.
FT	CONFLICT	1170	1170	/FTID=VAR_004078.
FT	CONFLICT	1170	1170	P -> A (IN REF. 2).
SEQ	SEQUENCE	1255 AA; 137909 MW; 3989DFDA04DC962 CRC64;		

Query Match 97.4%; Score 6640; DB 1; Length 1255;
 Best Local Similarity 97.5%; Pred. No. 0;
 Matches 1225; Conservative 8; Mismatches 21; Indels 2; Gaps 2;

QY	1	MELALCRWGLLLALPPGAASVCTGTDMKRLPASPEHLDMLRHLYOGCQVGNL	60
DB	1	MELALCRWGLLLALPPGAASVCTGTDMKRLPASPEHLDMLRHLYOGCQVGNL	60
QY	61	ELTYLPNTASISFLDIDIEVGVYLIHNOVRVPLRLIVRGTOLEFNVAVALVDNG	120
DB	61	ELTYLPNTASISFLDIDIEVGVYLIHNOVRVPLRLIVRGTOLEFNVAVALVDNG	120
QY	121	DLNNTTAVTASPGGLRELOLRSLTEILKGVLIQNNPOLCYDITLMDIPKNNOLA	180
DB	121	DLNNTTAVTASPGGLRELOLRSLTEILKGVLIQNNPOLCYDITLMDIPKNNOLA	180
QY	181	LTLIDTNRSRACHPCSPCKSGSRCKWSSSDCOSLRTVAGGACARCKGLPTDCHEQC	240
DB	181	LTLIDTNRSRACHPCSPCKSGSRCKWSSSDCOSLRTVAGGACARCKGLPTDCHEQC	240
QY	241	AAAGTGPMSDCLALPHNHSIGICELCPALVYNTDTFESMNPBGRVYFGASCYTAC	300
DB	241	AAAGTGPMSDCLALPHNHSIGICELCPALVYNTDTFESMNPBGRVYFGASCYTAC	300
QY	301	YNYLSTDVSGCTIYCPILNQEVYTRDGTQCEKSKRCACVAYCGLGKQYIKANSKITGLT	360
DB	301	YNYLSTDVSGCTIYCPILNQEVYTRDGTQCEKSKRCACVAYCGLGKQYIKANSKITGLT	360
QY	361	ELERAGCKTIFGSLAFPLPESFDGDPASNTAPLOPEQLOVETLEITGYLYISAMFDSLP	420
DB	361	ELERAGCKTIFGSLAFPLPESFDGDPASNTAPLOPEQLOVETLEITGYLYISAMFDSLP	420
QY	421	DLSPQNLQVIRKGIILNNGAYSLTLOGLISWGLRLSRLSGGLALIHNNTHLCFVHTV	480
DB	421	DLSPQNLQVIRKGIILNNGAYSLTLOGLISWGLRLSRLSGGLALIHNNTHLCFVHTV	480
QY	481	PMQDLFNPQALHTNRPEDDEVGGLACHOLCARGHGKMGSPQCNVCSQFLGQGC	540
DB	481	PMQDLFNPQALHTNRPEDDEVGGLACHOLCARGHGKMGSPQCNVCSQFLGQGC	540
QY	541	VEECRVYQGLPREVYNAHCLPCHPEQOPNGSVTCFGBADQCVACAHYKBPFCVAC	600
DB	541	VEECRVYQGLPREVYNAHCLPCHPEQOPNGSVTCFGBADQCVACAHYKBPFCVAC	600
QY	601	PSGVKPLSTMPIMKPFDEBGAQOPCINTGSCVLDKXGCAEDRASPLISYSAVVG	660
DB	601	PSGVKPLSTMPIMKPFDEBGAQOPCINTGSCVLDKXGCAEDRASPLISYSAVVG	660
QY	661	ILVVVVLGVVFGILIKRRQOKIRKTYMRRLQETELVEBPLTPSGAMPNQAQRILKEQYI	720
DB	661	ILVVVVLGVVFGILIKRRQOKIRKTYMRRLQETELVEBPLTPSGAMPNQAQRILKEQYI	720


```

FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138631 MW; 6129264583011402 CRC64;

```

```

Query Match 85.8%; Score 5849; DB 1; Length 1257;
Best Local Similarity 85.9%; Pred. No. 4,2e-309;
Matches 1081; Conservative 55; Mismatches 118; Indels 4; Gaps 4;

```

```

QY 1 MELALCRMGILLALLPPGAASOVCTGTDKRLRPASPEHLDMLRHLYOGGOVVOGNTL 60
DB 1 MELANCMWGLLALLPPGIACTGCTGDKRLRPASPEHLDMLRHLYOGGOVVOGNTL 60
QY 61 ELTYLPTNASLFLQDIOEVQGYVLIANQYQVPLQRLRIVRGTOLFEDNYALAVLDNG 120
DB 61 ELTYVPANASLFLQDIOEVQGYVLIANQYQVPLQRLRIVRGTOLFEDKXALAVLDNR 120
QY 121 DELNNTTPTV-GASGGIREQLRLSTLEILKGVLIQRNPQLCYQDTILMKD.FHQNOL 179
DB 121 DPODVAASTPQRTPEGLRELQRLSLTEILKGVLIQRNPQLCYQDVVLMKDVFRKNOL 180
QY 180 ALLTIDNRSRACHPCSPCKSGSRGWESSEDCSLTRTYCAGAGCARCKGLPTDCCHEO 239
DB 181 APVDIDNRSRACHPCAPACKDNHGWESPDCQILGTICTSGACRCKRLPTDCCHEO 240
QY 240 CAAGCTGPKHSDCLACHFNHSGICEHCPALVTYNTDTESHPNBEGRYTFGASCVTAC 299
DB 241 CAAGCTGPKHSDCLACHFNHSGICEHCPALVTYNTDTESHPNBEGRYTFGASCVTAC 300
QY 300 PNYLSTDVGSCTLVCPHNOEVAEDGTORCEKCSRPCARVCYGMQVYKANSKFEI 359
DB 301 PNYLSTEVGSCTLVCPHNOEVAEDGTORCEKCSRPCARVCYGMQVYKANSKFEI 360
QY 360 TELPACGCKIFGSLAPLPESFDGDPASNTAPLQPEQLQVFETLEITGYLYISAMPDL 419
DB 361 NVQEFDCGCKIFGSLAPLPESFDGDPASNTAPLQPEQLQVFETLEITGYLYISAMPDL 420
QY 420 PDLVFNOLQVIRGRILLHNGAYSLTLOGISWLGIRSLRELSSGALLHNTHLCPVHT 479
DB 421 RDLVFNOLQVIRGRILLHNGAYSLTLOGISWLGIRSLRELSSGALLHNTHLCPVHT 480
QY 480 VPMQDLFRNPHQALLHTANREDE-CVGEGLACHQIOLCARHCHGPGPTCCVNSQFLRQ 538
DB 481 VPMQDLFRNPHQALLHTANREDE-CVGEGLACHQIOLCARHCHGPGPTCCVNSQFLRQ 540
QY 539 ECVEECRVLOGLPREYNNAHCLPCHCECPQNGSVTCREPPADQVAAAHYKDPFCYA 598
DB 541 ECVEECRVWGLPREYVSDKRCPCHECCPQNSSECFSESDQCAAAHYKDSSCYA 600
QY 599 RCPGKVPDLSYMPINKFDEGACQPCPINTCHSCVLDLDDKCAPQASPLTSIVASV 658
DB 601 RCPGKVPDLSYMPINKFDEGACQPCPINTCHSCVLDLDDKCAPQASPLTSIVASV 660
QY 659 VGIILVVVLGVVFPILIKRQOQIKRYTWARLLQETELVEPLTPSGAMPNOQAMRLKEQ 718
DB 661 VGIILVVVLGVVFPILIKRQOQIKRYTWARLLQETELVEPLTPSGAMPNOQAMRLKET 720
QY 719 YIKANSKFIGITEL-TVYKGIWIPDGENVYKIPVAIVLRENTSPKANKRILBEAVYMGV 777
DB 721 ELR-KYVVLGSGAGFYTKGIWIPDGENVYKIPVAIVLRENTSPKANKRILBEAVYMGV 779
QY 778 GSPYVSRLLGICLTSTVQVLTQMLPFQCLLDHVENRGLSGODLLMCMQIAKQSYLE 837

```

```

DB 780 GSPYVSRLLGICLTSTVQVLTQMLPFQCLLDHVENRGLSGODLLMCMQIAKQSYLE 839
QY 838 DVLRLVHDLAARNLVKSPNVRKITTDEGLARLLDIDETEVHAGQKVPKIMMALESILRR 897
DB 840 DVLRLVHDLAARNLVKSPNVRKITTDEGLARLLDIDETEVHAGQKVPKIMMALESILRR 899
QY 898 RFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVYIMYK 957
DB 899 RFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVYIMYK 959
QY 958 CMWTDSECRPFRETLVEFSRMAPDRPFVIONEDLGPASPLDSFFYSLEDDMDGL 1017
DB 960 CMWTDSECRPFRETLVEFSRMAPDRPFVIONEDLGPASPLDSFFYSLEDDMDGL 1019
QY 1018 VDABEYLVPOQGFPCPDPAQAGQVWHRSSSTRSGGDLTLGLEPSEEAAPSPLAP 1077
DB 1020 VDABEYLVPOQGFSPDPTOTGSTAHRHRSSSTRSGGDLTLGLEPSEEAAPSPLAP 1079
QY 1078 SEGAGSDVFDGDLGMGAAXGLQSLPTHDESPLORYSEDPVPLPSETDGVAPLTCSPQ 1137
DB 1080 SEGAGSDVFDGDLGMGYTKGLQSLPHDLSPLQRYSEDPVPLPSETDGVAPLTCSPQ 1139
QY 1138 EYVNAQPVRRPQPSRPGPLPAARPAQATLERAKTUSPGKNGVYKQVFAFGAVENPEYL 1197
DB 1140 EYVNAQSVQPPPLTPGGLPVPYRPAQATLERAKTUSPGKNGVYKQVFAFGAVENPEYL 1199
QY 1198 TPQCGAAPQPPPPAPSPAFDNTLYMDQDPERGAPSPSTFKGTAEENEYGLDVPV 1255
DB 1200 VPREGTASPPHSPAPSPAFDNTLYMDQDSEGGPPSPNEGTAEENEYGLDVPV 1257

RESULT 3
ERR2_MESAU
ID ERR2_MESAU STANDARD; PRT; 1254 AA.
AC Q60553;
DT 15-DEC-1998 (Rel. 37, Created)
DR 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor Protein-tyrosine kinase etdb-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (neu proto-oncogene) (c-erbB-2).
GN ERBB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
K1 [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94193007; Pubmed=7908275;
RA Nakamura T., Ushijima T., Ishikawa Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.,
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255 (1994).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PPM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 or send an email to license@isb-sib.ch.

DR EMBL: D16295; BAA03801.1; -
 DR HSSP: P11362; 1PCK.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR Pfam: PF02757; YLP; 2.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; FU; 3.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transmembrane; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Proto-oncogene; Disease mutation.
 FT SIGNAL 1 21
 FT CHAIN 1 21
 FT DOMAIN 22 1254
 FT TRANSMEM 652 652
 FT DOMAIN 676 1254
 FT DOMAIN 158 368
 FT DOMAIN 472 644
 FT DOMAIN 720 987
 FT NP_BIND 726 734
 FT BINDING 753 753
 FT ACT_SITE 845 845
 FT DISULFID 195 204
 FT DISULFID 199 212
 FT DISULFID 236 244
 FT DISULFID 240 252
 FT DISULFID 255 264
 FT DISULFID 268 295
 FT DISULFID 299 311
 FT DISULFID 315 331
 FT DISULFID 334 338
 FT DISULFID 511 520
 FT DISULFID 515 528
 FT DISULFID 531 540
 FT DISULFID 544 560
 FT DISULFID 563 576
 FT DISULFID 567 584
 FT DISULFID 587 596
 FT DISULFID 600 623
 FT DISULFID 626 634
 FT DISULFID 630 642
 FT MOD_RES 1139 1139
 FT MOD_RES 1247 1247
 FT CARBOHYD 68 68
 FT CARBOHYD 125 125
 FT CARBOHYD 187 187
 FT CARBOHYD 259 259
 FT CARBOHYD 530 530
 FT CARBOHYD 571 571
 FT CARBOHYD 629 629
 FT CARBOHYD 658 658
 FT VARIANT 659 659
 FT VARIANT 659 659
 FT SEQUENCE 1254 AA; 138252 MW; 974C3791C31F2BE1 CMC64;

Query Match 85.6%; Score 5833.5; DB 1; Length 1254;
 Best Local Similarity 85.5%; Pred. No. 2.9e-308;
 Matches 1074; Conservative 64; Mismatches 115; Indels 3; Gaps 3;

QY 1 MELALCRWGLLALLPFGAASSTVCTGTDMLRLPASPETHLDMRLHYOGGVQVQNTL 60
 Db 1 MELAAWGMGLLALLSPGASSTVCTGTDMLRLPASPETHLDMRLHYOGGVQVQNTL 60

QY 61 ELTYLPTNASLSEFLDIOIQGVQVULIANVQRVPLQRLIVRGTLQFEDNYALALVDNG 120
 Db 61 ELTYLPANVTLSLFLDIOEQGYMILIAHSVRHPLQLRLIVRGTLQFEDNYALALVDNR 120
 QY 121 DPLNNTPTVPGASPGALRELOLRSLTELKGVGLIQRVPOCYOQPTIMKDIPIKNNOLA 180
 Db 121 DPLDNTNTTGTTPGLRELOLRSLTELKGVGLIRKVPQCLQDTIMKQVFRKNNOLA 180
 QY 181 LTLIDTNRSPACHPSPMCKSGRCWSESSDCCSLTRTVACGACRCKGRLPTDCHEQC 240
 Db 181 PYDIDTNRSPACPCAPACKDNHCWGAPEPCQTLITGIABRAVAPAAARLPTDCHEQC 240
 QY 241 AAGCTGPKKSDCLACHTNHSICILHCPALVTYNTDFFESMPREGRYTGGACVYACP 300
 Db 241 AAGCTGPKKSDCLACHTNHSICILHCPALVTYNTDFFESMPREGRYTGGACVYTCP 300
 QY 301 YNYLSTDVGSCTLVCPLANOEVTAEADGTORCEKSKPCARVCYGLGMQYIKANSKFIGIT 360
 Db 301 YNYLSTEVGSCTLVCPPLNNOEVTAEADGTORCEKSKSGARVCYGLGMHRLGALAITSAN 360
 QY 361 ELFFAGCKKIFGSLAFLEPESDGPASNTAPLOEPOLOVFETLEITGYLISAMPDLP 420
 Db 361 IOEFAGCKKIFGSLAFLEPESDGPASNTAPLOEPOLOVFETLEITGYLISAMPDLP 420
 QY 421 DLSVPLQNLQVIRGRILHNGAVSLTLOGIGIWMGLRSRLREGSLALIHNTTLCFVHTV 480
 Db 421 DLSVQNLQVIRGRILHNGAVSLTLOGIGIWMGLRSRLREGSLALIHNTTLCFVHTV 480
 QY 481 PWDQLEPNNPHALHTNANPEDECVGEGLAGHOLCARHCWGPPTCVNCSQFLRQEC 540
 Db 481 PWDQLEPNNPHALHTNANPEDECVGEGLAGHOLCARHCWGPPTCVNCSQFLRQEC 540
 QY 541 VEEGVTLGLREYVYNAHCLPCPEEQPQNGASTCCGPEPDQCVAAHYKDPFCVARC 600
 Db 541 VEEGVTLGLREYVYNAHCLPCPEEQPQNGASTCCGPEPDQCVAAHYKDPFCVARC 600
 QY 541 VEEGVTLGLREYVYNAHCLPCPEEQPQNGASTCCGPEPDQCVAAHYKDPFCVARC 600
 Db 541 VEEGVTLGLREYVYNAHCLPCPEEQPQNGASTCCGPEPDQCVAAHYKDPFCVARC 600
 QY 601 PSQVFPDLSYMPIMKFPDEBGACQPCFINTCHSCVDLDDGCPAEQASPLTSIVAVG 660
 Db 601 PSQVFPDLSYMPIMKFPDEBGACQPCFINTCHSCVDLDDGCPAEQASPLTSIVAVG 660
 QY 661 ILVVVGLGVGILIKRQOKIRKTYMRLLQETLEVEPLTPSGAMPNQMRILKEQYI 720
 Db 661 ILVVVGLGVGILIKRQOKIRKTYMRLLQETLEVEPLTPSGAMPNQMRILKEQYI 720
 QY 721 KANSKFIGITEL-TVYKGIWIPQGENVKIPVAKIVLBENTSPRANKELIDEAVVMGVGS 779
 Db 721 R-KYKVLSSGAFGVYVYKGIWIPQGENVKIPVAKIVLBENTSPRANKELIDEAVVMGVGS 779
 QY 780 PYVSRLLGICLTSTVQCVTOLMPYGCILLDHVRENGRLSGQDLLNMCQIAKQMSYLEDV 839
 Db 780 PYVSRLLGICLTSTVQCVTOLMPYGCILLDHVRENGRLSGQDLLNMCQIAKQMSYLEDV 839
 QY 840 RLYVRDLAARVAVLKSNHVKITDFFGLARLLDIDETEHADGKVPKMALESLIRRF 899
 Db 840 RLYVRDLAARVAVLKSNHVKITDFFGLARLLDIDETEHADGKVPKMALESLIRRF 899
 QY 900 THQSDVMSYGVYVWELMTFGAKPYDGIIPASEIPDLLEKGERLPPPICTIDVYMIWVCM 959
 Db 900 THQSDVMSYGVYVWELMTFGAKPYDGIIPASEIPDLLEKGERLPPPICTIDVYMIWVCM 959
 QY 960 MIDSECPREELVSESSRMAKPPORFVYI QNEDLPASFLDSTPYRSLLEDMDMDLVD 1019
 Db 960 MIDSECPREELVSESSRMAKPPORFVYI QNEDLPASFLDSTPYRSLLEDMDMDLVD 1019
 QY 1020 ABEYTLVPOQGFPCPDPAAGAWHHRHSSSTRSGGDLTGLFESEEARSPPLAPSE 1079
 Db 1020 ABEYTLVPOQGFPCPDPAAGAWHHRHSSSTRSGGDLTGLFESEEARSPPLAPSE 1079
 QY 1080 GAGSDVVDQDLGMAAKGLQSLPTHPSPQLQRYSEDPTVPLPSETDGYVAPYLTCSPQPEY 1139
 Db 1080 GAGSDVVDQDLGMAAKGLQSLPTHPSPQLQRYSEDPTVPLPSETDGYVAPYLTCSPQPEY 1139

```

QY      1140 VNQDVRPPPPPPRGDPLPAAPACATIERAKTISPGKNGVYKQVFAFGAVENPEVLT 1199
DB      1140 VNQPEVRPPPLTPGCPPEPAPACATIERKPTISPGNGVYKQVFTFGAVENPEVLT 1199
QY      1200 OCGAAPPPPPPPAFPDNLYYMDQDPPERCAPSTFKGTPTAENPEYGLDVPV 1255
DB      1200 RGGASQPH-PPALCPARFDNLYYMDQDPSERKSPNTEGTPTAENPEYGLDVPV 1254

RESULT 4
EGFR HUMAN STANDARD: PRT: 1210 AA.
ID EGFR_HUMAN P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92735; Q00732;
AC P00573; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92735; Q00732;
AC Q00698; Q9B2S2; Q9H2C9; Q9G2X1; Q9H3C9;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
DE protein-tyrosine kinase ErbB-1).
GN EGFR OR ERBB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RX SEQUENCE FROM N.A. (ISOFORM 1).
RA MEDLINE=84219729; PubMed=6328312;
RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
RA Lee Y., Yarden Y., Libermann T.A., Schlessinger U., Downward J.,
RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
RT "Human epidermal growth factor receptor cDNA sequence and aberrant
RT expression of the amplified gene in A431 epidermoid carcinoma cells.";
RL Nature 309:418-425 (1984).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX TISSUE=Placenta; PubMed=7654368;
RA Ilekis J.V., Stark B.C., Scoccia B.;
RT "Possible role of variant RNA transcripts in the regulation of
RT epidermal growth factor receptor expression in human placenta.";
RL Mol. Reprod. Dev. 41:149-156 (1995).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX TISSUE=Placenta; PubMed=8918811;
RA Reiter J.L., Mahlie N.U.;
RT "A 1.8 kb alternative transcript from the human epidermal growth
RT factor receptor gene encodes a truncated form of the receptor.";
RL Nucleic Acids Res. 24:4050-4056 (1996).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX TISSUE=Placenta; PubMed=9103388;
RA MEDLINE=97256547; PubMed=9103388;
RA Ilekis J.V., Garliti J., Niederberger C., Scoccia B.;
RT "Expression of a truncated epidermal growth factor receptor-like
RT protein (TEGFR) in ovarian cancer.";
RL Gynecol. Oncol. 65:36-41 (1997).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
RX TISSUE=Placenta; PubMed=1161793;
RA MEDLINE=21100872; PubMed=1161793;
RA Reiter J.L., Thredgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Scheil Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampand A.L.,
RA Balasubramanian S., Crossley T.D., Magnuson T.R., James C.D.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative EGFR transcripts encoding truncated receptor
RT isoforms.";
RL Genomics 71:1-20 (2001).
RN [6]
RP SEQUENCE OF 575-687 FROM N.A.
RA Reiter J.L., Thredgill D.W., Danielsen A.J., Scheil C.M.,
RA Lampand A.L., Balasubramanian S., Crossley T.O., Magnuson T.R.,

```

```

RA Mahlie N.U.;
RT "Human and mouse alternative EGFR transcripts encoding only the
RT extracellular domain of the receptor.";
RL Submitted (Feb-1999) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 713-924 FROM N.A.
RX MEDLINE=84196372; PubMed=6326261;
RA Lin C.R., Chen W.S., Krueger W., Stolarek L.S., Weber W.,
RA Evans R.M., Verma I.M., Gill G.N., Rosenfield M.G.;
RT "Expression cloning of human EGF receptor complementary DNA: gene
RT amplification and three related messenger RNA products in A431
RT cells.";
RL Science 224:843-848 (1984).
RN [8]
RP SEQUENCE OF 150-962 FROM N.A.
RX MEDLINE=84245835; PubMed=6330563;
RA Xu Y.H., Ishii S., Clark A.J., Sullivan M., Wilson R.K., Ma D.P.,
RA Roe B.A., Merlino G.T., Pastan I.;
RT "Human epidermal growth factor receptor cDNA is homologous to a
RT variety of RNAs overproduced in A431 carcinoma cells.";
RL Nature 309:806-810 (1984).
RN [9]
RP SEQUENCE OF 1028-1210 FROM N.A.
RX MEDLINE=85046483; PubMed=6093780;
RA Simmen F.A., Gope M.L., Schultz T.Z., Wright D.A., Carpenter G.,
RA O'Malley B.W.;
RT "Isolation of an evolutionarily conserved epidermal growth factor
RT receptor cDNA from human A431 carcinoma cells.";
RL Biochem. Biophys. Res. Commun. 124:125-132 (1984).
RN [10]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=88217333; PubMed=3329716;
RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
RA Waterfield M.D.;
RT "The human EGF receptor gene: structure of the 110 kb locus and
RT identification of sequences regulating its transcription.";
RL Oncogene Res. 1:375-396 (1987).
RN [11]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=91107677; PubMed=1988448;
RA Haley J.D., Waterfield M.D.;
RT "Contributory effects of de novo transcription and premature
RT transcript termination in the regulation of human epidermal growth
RT factor receptor proto-oncogene RNA synthesis.";
RL J. Biol. Chem. 266:1746-1753 (1991).
RN [12]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=85270438; PubMed=2991899;
RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
RT "Characterization and sequence of the promoter region of the human
RT epidermal growth factor receptor gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924 (1985).
RN [13]
RP SEQUENCE OF 540.
RA Kohda D.;
RT Submitted (SEP-1997) to the SWISS-PROT data bank.
RN [14]
RP RECEPTOR ACTIVITY.
RX MEDLINE=84191554; PubMed=6325948;
RA Mroczkowski B., Mosig G., Cohen S.;
RT "ATP-stimulated interaction between epidermal growth factor receptor
RT and supercoiled DNA.";
RL Nature 309:270-273 (1984).
RN [15]
RP PHOSPHORYLATION.
RX MEDLINE=89278137; PubMed=2543678;
RA Margolis B.L., Lax I., Kris R., Domagalian M., Hongger A.M.,
RA Howk R., Givol D., Ullrich A., Schlessinger U.;
RT "All autophosphorylation sites of epidermal growth factor (EGF)
RT receptor and HER2/neu are located in their carboxyl-terminal tails.
RT Identification of a novel site in EGF receptor.";
RL J. Biol. Chem. 264:10667-10671 (1989).
RN [16]

```


RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
 RP ASN-528.
 RX MEDLINE=96398132; PubMed=8962717;
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
 RT "Analysis of the glycosylation patterns of the extracellular domain of
 the epidermal growth factor receptor expressed in Chinese hamster
 ovary fibroblasts.";
 RL Growth Factors 13:121-132(1996).
 RN [17]
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
 RP ASN-603.
 RX MEDLINE=20198209; PubMed=10731668;
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
 RT "Characterization of the N-oligosaccharides attached to the atypical
 Asn-X-Cys sequence of recombinant human epidermal growth factor
 receptor.";
 RL J. Biochem. 127:65-72(2000).
 RN [18]
 RP PARTIAL SEQUENCE AND DISULFIDE BONDS.
 RX MEDLINE=98225196; PubMed=9556602;
 RA Abe Y., Oda M., Inagaki F., Iax I., Schlessinger J., Kohda D.;
 RT "Disulfide bond structure of human epidermal growth factor receptor.";
 RL J. Biol. Chem. 273:11150-11157(1998).
 RN [19]
 RP REVIEW
 RX MEDLINE=87297456; PubMed=3039909;
 RA Carpenter G.;
 RT "Receptors for epidermal growth factor and other polypeptide
 mitogens.";
 RL Annu. Rev. Biochem. 56:881-914(1987).
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF
 family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
 EGF-like growth factor, Gp30 and vaccinia virus growth factor. Is
 involved in the control of cell growth and differentiation.
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
 secreted.
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/
 truncated isoform/TEGFR, 3/p110 and 4; are produced by
 CC alternative splicing.
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
 expressed in ovarian cancers.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 dimerization, internalization of the EGF-receptor complex,
 induction of the tyrosine kinase activity, stimulation of cell DNA
 synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: X00588; CA035240.1; -
 DR EMBL: U95088; AAB53063.1; -
 DR EMBL: U48722; AAC50802.1; -
 DR EMBL: U48723; AAC50804.1; -
 DR EMBL: U48724; AAC50796.1; -
 DR EMBL: U48725; AAC50797.1; -
 DR EMBL: U48726; AAC50798.1; -
 DR
 Query Match 45.0%; Score 3065; DB 1; Length 1210;
 Best Local Similarity 49.2%; Pred. No. 1,6e-158;
 Matches 623; Conservative 173; Mismatches 361; Indels 110; Gaps 25;

QY 69 ASLSFLDIDQEGGVYVLIANNOVROVPLQRLATVRGTOLFEVYVALVLDNGPLNNTTP 128
 DB 74 YDLSEFLKTIQEVAVYVLIANTVERPLENIQIRGMYYENSVALAVSNVD----- 126
 QY 129 VTGASPGELRELQSRSTELIKSGVLIQRPOCYDPTLMKDIPIKNNQALATLIDTR 188
 DB 127 ---ANKTGLEIPKRNQELIHGAVFSSNPALCNVESTIOMRIVSSDFLSNMSKPFONH 183
 QY 189 SRACHSPMCCKSRKCSWGESEEDCGSLTRVTCAGCA-ICKGPLEPTDCHQCAAGCTGP 247
 DB 184 LGSQCKDPSCPNWSCWGAEEFNCQKLTIKICAQQCSGRCKRSPSDCCHNQCAAGCTGP 243
 QY 248 KHSPCLACFHNNISGICELCPALVYNTDFTESMNPGRRTYFGASCTYACPYNTLSD 307
 DB 244 RESDCLVRKFRDQATCKDCCPLMLYNFTYQMDVNPGEKYSFGATCYKCPERNVYVD 303
 QY 308 VGSCTLVCPLNQEVTAEDGTQRCCKSPRCARVCGYGLMOYIKANSKFIGTELE-FAG 366
 DB 304 HGSQVRACGADSYEM-EDGVKRCCKCEGRCRVKNGIDIGERK-DLSLNATNIGHFKK 361
 QY 367 CKIFGSLAPSPSFDGDPASNTAPLQPELOVFEFLBITGYLYISAPDSLPLDSVFQ 426
 DB 362 CTISGDLHLPLVAFRGDSFTHTPPLDPELDILKTVKEITGFLIQAMPENRTDLHAF 421
 QY 427 NLQVIRGRILHNGAVSLTLOGIGISMLGRSLRELGSLALIHNNHTLGFVHTVPDQLF 486
 DB 422 NLEIRKRTQHQGFSLAVYSLNTSLGRSLKEISDGVYISGNKNLCYANTINMKLIF 481
 QY 487 RNPHQALHTANRDEDECVGEGSLAQHQLCARGCMGPGPTQCVNCSQFLRGCEVECRV 546
 DB 482 GTSGQCKTKIISNGENSCAKATGQVCHALCSPGCQWPEPRDCVSNVSRGECVDKCL 541
 QY 547 LOGIPREYVNAARCLCHECEQPNQNSVTCFREPADQVCAAHYDPPPCVAKRSGVKP 606
 DB 542 LEBEPREFENSCFICHPCELPQAMNITCTGRPNCTIQCAHYDIPGRCVATCPAGVWG 601
 QY 607 DLSYPMKRPDEEGACOPCPINCTHSVDLDDKGPAPQASPLTSIYSAVG---ITL 663
 DB 602 ENNTL-VMKXADAGHYCHCHRNCTYCGTGPGEPTNGPRIP--STATGVAGALLLL 658
 QY 664 VVTLGVVFGILIRROOKRKTYMRRLIETELVPLTPSGAMPQAWRIKEQYIRAN 723
 DB 659 VVALGIG--LEMRRHIVKRTLRLLQERELVPLTPSGAPQALLRIKETEFK-K 714
 QY 724 SKPIGITEU-TYKGIWIDGSEVVKIPVATKYLRENTSPKAKETLDAVYVWAGSGPYV 782
 DB 715 IKVLGSGAGTYVKGMLPEGEKVKIPVATKELRENTSPKAKETLDAVYVWAGSDNPHV 774
 QY 783 SRLIGICLTSTVQVLTQMLPYGCLLDHYRENGRLGSDLLNMCQIAKSGSYLEDVLY 842
 DB 775 CRLIGICLTSTVQVLTQMLPYGCLLDHYREHNDNGSYLLNMCQIAKSGSYLEDRLV 834
 QY 843 HRLAARNTLVSPNHYKTTDGLARLIDIDTEHNAAGKVPITMMALLESILRRRTFHQ 902
 DB 835 HRLAARNTLVTPQVKTITDGLAKLGAEEKEVHAAGKVPITMMALLESILHRTYHQ 894
 QY 903 SDVMSYGVTVMEIHTFGAKPYDIPAREIPDLLEKGERLPORPCTIVYVIMYKCMWID 962
 DB 895 SDVMSYGVTVMEIHTFGSKPYDIPAREIPDLLEKGERLPORPCTIVYVIMYKCMWID 954
 QY 963 SECRPFRELVSEFSRMAADPQRFVYVIO-NEDLGPASPLSTFYKSLLEDDMDMDVDAE 1021
 DB 955 ADSRPFRELVSEFSRMAADPQRFVYVIO-NEDLGPASPLSTFYKSLLEDDMDMDVDAE 1014
 QY 1022 EYLVPOQGFPCDPAFAGAGVNHHRSSSRSGGDLTLLEPSBEBAAPSLAPSEGA 1091
 DB 1015 EYLVPOQGF-----SSPSTSTPTLLSLISA 1040
 QY 1082 GSDVVDGDIQGAARGLQSLPTHDSPLQARYSEDPVPLPSET--DGVAFLTQSPQPEY 1139
 DB 1041 TSN--NSTVACIDRNGLSQSPFKEDSFQRYSSDPTALTDSIDDTFL-----PVPEY 1092


```

FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY)
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY)
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE)
FT CARBOHYD 128 128 (BY SIMILARITY)
FT CARBOHYD 175 175 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 196 196 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 352 352 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 413 413 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 444 444 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 528 528 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 568 568 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 603 603 N-LINKED (GLCNAC) (POTENTIAL)
FT CARBOHYD 623 623 N-LINKED (GLCNAC) (POTENTIAL)
FT CONFLICT 19 19 C -> S (IN REF. 2)
FT CONFLICT 539 539 C -> W (IN REF. 5)
FT CONFLICT 991 991 L -> F (IN REF. 4)
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6)
SQ SEQUENCE 1210 AA, 134853 MW, 690E20D46DF2D2F5 CRC64;

```

```

Query Match 44.6%; Score 3039; DB 1; Length 1210;
Best Local Similarity 49.1%; Pred. No. 4,2e-157;
Matches 625; Conservative 167; Mismatches 368; Indels 114; Gaps 27;

```

```

QY 11 LLLALLPQAA--STQVCTGTDMLRLPASPEHLDMRLHYOGQVYQNTLEITYPTN 68
DB 14 LTLALCAAGALBKKVCGGTNRLLTQCTPEDHLSLQRYNNCEVGLNLEITYQNRN 73
QY 69 ASLSFLQDIQEVQGVYLAHNOVRVPLQRLRYVGTQLFEDNALVALVDNGDLNNTTP 128
DB 74 YDLSFLKTIQEVAGVYLAHNTVERIPLNLOIIRGNALYENTYALALISN----- 124
QY 129 VTGASPGGLREQLSLREILKGVLIQRNPQLCQDTIILMKDI-----FHQNOALTLI 184
DB 125 -GTRRTGLRELPMNLQELILGAVRSNNPILCMWDITQMRDVIQVNVFMSNMSMDI--- 180
QY 185 DTNRSRACHPCSPMKSGSRGWSESEDCSLTRTVACAGCA-RCKGFLPTDCHQCGAAG 243
DB 181 -QSHPSGPCDPCSPGSCGSGGEGNCQKLTIIICQCGHRGRGSPDCHCNCQAG 229
QY 244 CTGPKHSCLACLNHNHSGICEIHPALVTNTPTFESMPREGRYFGASCYTACPYNY 303
DB 240 CTGPRESOLVQKQDEBATCKDTCPPMLNPTTYQMDVNPBEKYSFGATCYVKKCPRNY 299
QY 304 LSTDVGSCTLYCPILHNOEVTAEADGTORCEKSKPCARCVGLGMQYIKANSKFGITELE 363
DB 300 VYTDHGSCTVACGPGYEV-EDGIRKKKDDGCRKVCNGIGIEPK-DTSLINATNIK 357
QY 364 -FAGCKKIFGSLAFPESEFDGPPASNTAPLQPEQLQVFETLEITGYLYISAMPDLPDL 422
DB 358 HFXYCTAISGDLHLPLVAFKGDSTRTPTPLPRELEILKTVKEITGFLIQAMPDMWTDL 417
QY 423 SVFQNLQVIRGRILJHNGAVSLTLQGLISWLSASLELQSLALIHNNHLCFVNTIVM 482
DB 418 HAFENLEIRGRITQHQOFLAVGLVNTISLGLSLKEISDGDVILISGNNILCYANTIVM 477
QY 483 DQLEPNHQALLHTANPREDECEVGEGLACHQLCARHGWSPGFTQCVNCSQPLRQGEVE 542
DB 478 KKLFGTPQOKTKIMNNAEKCKAVNHNVPILCSSEGCWGPBEPDVCSCQNSRGKECIE 537
QY 543 ECRVLOGLPREYVARRHCLPCHPCOPONGSVTCFGEALQCCACAHYKDPPTCVARCS 602
DB 538 KGNILGEPRESEFVNSECIQHPCLPQAMNITCTTGPNQNCIQCAHYIDGPRCVATCPA 597
QY 603 GVKPDLSTMPWKPEDEGACQPCPINCSTHSCVDLDKGPAPACORASPLTSISAVVGL 662
DB 598 GIMGENNTIL-VMKYADANNVCHLCHACTYGCAGPGLQGEVWSPSPKIPSTATIGVGL 656
QY 663 LVVVLGVVFGI-LIKRQOKIRKTYTKRRLIQETELVEPLTPSSGAMPNQOMRLKEQYIK 721
DB 657 LFTIV-VALGIGLWRRRHIVRKRTLRRLIQERELVEPLTPSGEAPNOAHLILKETEFK 715
QY 722 ANSKFGITEL-TVYKGIWIPDGENVKIPVAIKVLBNSTSPKXNKELIDEAVYMGVGP 760

```

```

DB 716 -KIKVLGSGAFGVYKGLIPEGEVKIPVAIKELRATSPRANKELIDEAVYMSVNP 774
QY 781 YSRLLIGLITSTVQLVQLMYGGLDHNENRGRLOSLDLAMCQIANGMYLEDDR 840
DB 775 HVCRLIGLITSTVQLVQLMYGGLDHNENRGRLOSLDLAMCQIANGMYLEDDR 834
QY 841 LVHRDLAANVLVYKPNHYKIDFGLARLDIDENYADGKVPKMALESLIRPFT 900
DB 835 LVHRDLAANVLVYKPNHYKIDFGLARLDIDENYADGKVPKMALESLIRPFT 894
QY 901 HOSDWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPEPCTIDYMIWVCM 960
DB 895 HOSDWSYGVTVWELMTFGSKPYDGIPIASDISILEKGERLPPEPCTIDYMIWVCM 954
QY 961 IDSECRPRELIVSEFPMARPPQPFVITQ-NEDGAPSPDLSTFYRLLEDMDGLVD 1015
DB 955 IDADSPKRELILFESKXADPPQYVLIQDDEHMLPSPTDSNFYALMDEEDMEDVVD 1014
QY 1020 AEEYLVPOQGFPCPPAPAGAGMVRHRSSTSGGDLTLGLEPSEAPRSPAPSE 1075
DB 1015 ADEYVLIPOQGF-----NSPST-----SRTPILSL 1040
QY 1080 GAGSDVFGDLGMAKGLQSLPHDSEFLQRYSDPTVPLPST--DGVAAPLTCSPOP 1137
DB 1041 SATSN-----NSTVACINNGSCRYVEDAFLORYSDPTGAVTEDNIDAFI-----FVP 1090
QY 1138 EYVNPQVAPQPPSPREGPLPAARAGATLERATLSGKXGVVQVDFAFAGAVENEYL 1197
DB 1091 EYVNPQVAPQPPSPREGPLPAARAGATLERATLSGKXGVVQVDFAFAGAVENEYL 1139
QY 1198 -TPQGAAPQPPPPAPSPAFDNTYWDQ-----DP-----PERGAPSTFKGT 1240
DB 1140 NTAQ-----PCLSSGFSNPAIMQKSHQMSLDNDPQDFFPKETKRGIRKG- 1189
QY 1241 PTAENPEYLGIVP 1254
DB 1190 PTAENPEYLRVAP 1203

```

```

RESULT 6
ERR4_HUMAN
ID ERR4_HUMAN STANDARD; PRT; 1308 AA.
AC Q15303;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
DE (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
GN ERB4 OR HER4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM JM-A).
RC TISSUE=Breast carcinoma;
RX MEDLINE=93189574; PubMed=8383326;
RA Plowman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlson G.W.,
RA Foy L., Neubauer M.G., Shoyab M.,
RT "Ligand-specific activation of HER4/p180erbB4, a fourth member of the
RT epidermal growth factor receptor family".
RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
RC TISSUE=fetal brain;
RX MEDLINE=97476287; PubMed=9334263;
RA Elenius K., Cortas G., Paul S., Choi C.J., Rio C., Plowman G.D.,
RA Klagsbrun M.,
RT "A novel juxtamembrane domain isoform of HER4/erbB4. Isoform-specific
RT tissue distribution and differential processing in response to
RT phorbol ester."
RL J. Biol. Chem. 272:26761-26768(1997).

```

CC - FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
 CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
 CC NTAR. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
 CC NOT ACTIVATED BY EGF, TGF- α , AND AMPHIREGULIN.
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC - SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
 CC RECEPTORS (POTENTIAL).
 CC - SUBCELLULAR LOCATION: 1. TYPE I membrane protein.
 CC - ALTERNATIVE PRODUCTS: 2. ISOFORMS: JM-A (SHOWN HERE) AND JM-B;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
 CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
 CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
 CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
 CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
 CC - TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
 CC KIDNEY. IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,
 CC PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,
 CC LUNG, SALIVARY GLAND, AND PANCREAS.
 CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES.
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).

DR EMBL: L07868; AAB59446.1; -
 DR HSSP: P11362; 1FGK.
 DR Genew: HGNC:3432; ERBB4.
 DR MIM: 600543; -
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_kinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR Pfam: PF02757; YLP; 2.
 DR Prodom: PD000001; Euk_kinase; 1.
 DR SMART: SM00263; EU_4.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane: Glycoprotein; Multigene family; Receptor; Signal;
 KM Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KM Alternative splicing.
 FT SIGNAL 1 25
 FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
 FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 652 675 POTENTIAL.
 FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 186 334 CYS-RICH.
 FT DOMAIN 496 633 CYS-RICH.
 FT DOMAIN 718 985 PROTEIN KINASE.
 FT NP_BIND 724 732 ATP (BY SIMILARITY).
 FT BINDING 751 751 ATP (BY SIMILARITY).
 FT ACT_SITE 843 843
 FT DISULFID 189 197 BY SIMILARITY.
 FT DISULFID 193 205 BY SIMILARITY.
 FT DISULFID 213 221 BY SIMILARITY.
 FT DISULFID 217 229 BY SIMILARITY.
 FT DISULFID 230 238 BY SIMILARITY.
 FT DISULFID 234 246 BY SIMILARITY.
 FT DISULFID 249 258 BY SIMILARITY.
 FT DISULFID 262 289 BY SIMILARITY.
 FT DISULFID 293 304 BY SIMILARITY.

FT DISULFID 308 323 BY SIMILARITY.
 FT DISULFID 326 330 BY SIMILARITY.
 FT DISULFID 303 312 BY SIMILARITY.
 FT DISULFID 507 520 BY SIMILARITY.
 FT DISULFID 523 532 BY SIMILARITY.
 FT DISULFID 536 552 BY SIMILARITY.
 FT DISULFID 555 569 BY SIMILARITY.
 FT DISULFID 559 577 BY SIMILARITY.
 FT DISULFID 580 589 BY SIMILARITY.
 FT DISULFID 593 614 BY SIMILARITY.
 FT DISULFID 617 625 BY SIMILARITY.
 FT DISULFID 621 633 BY SIMILARITY.
 FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1168 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1258 1268 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 138 138
 FT CARBOHYD 174 174 N-LINKED (GLCNAC).
 FT CARBOHYD 181 181 N-LINKED (GLCNAC).
 FT CARBOHYD 253 253 N-LINKED (GLCNAC).
 FT CARBOHYD 358 358 N-LINKED (GLCNAC).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC).
 FT CARBOHYD 473 473 N-LINKED (GLCNAC).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC).
 FT CARBOHYD 548 548 N-LINKED (GLCNAC).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC).
 FT CARBOHYD 620 620 N-LINKED (GLCNAC).
 FT VARSPLIC 626 648 NGPISHCIYPMYGHSTLQHA -> IGSSIEDICGLMD
 (IN ISOFORM JM-B).
 SQ SEQUENCE 1308 AA; 146807 MW; 5E4A809B5D88761 CRC64;
 Query Match 42.4%; Score 2887.5; DB 1; Length 1308;
 Best Local Similarity 44.5%; Pred. No. 7.1e-149;
 Matches 599; Conservative 184; Mismatches 389; Indels 175; Gaps 30;
 9 WGLLALLPPGAA-----STVCTGTMKRLPASPETHLMLRLHYGCGVVOGNNLELY 64
 8 WWSVLVAAGTVPPSDSQSVCACTEKLSTSLDEQYALKRYENCCEVMGNLEITS 67
 65 LPTNASLSFLDIOEVOGVYLAHQVROVPLQBLVRGQVFPEDYALAVLDNDPUN 124
 68 IEHRDLSFLSVSEVTVYVALNORRYPLEBLRIRIGTKLYEDRYALATILNRKOG 127
 125 NTPYTGASPGQLRELOLSLTLEILKGVLIQRNPOLCYDITLMDIFKHNQALTLI 184
 128 NF-----GLOELGLKNTLEILNGGVYVDQKFKLYADTIHQDVIANPWSNLTIV 178
 185 DITNSRACHPCSPKCKSGESSEDCQSLTRVYVAGGC-ARCKGLPTDCCHEQCAAG 243
 179 STNGSSGGRCHKSCGTG-RCWGPENHCQTLTRVCAEQCDGRCYGVYSDCCHEQCAAG 237
 244 CTGPGHSDCLALHFHNSGICELCPALVTYNTDTEFSEMPREGRYFGASVYACPYNY 303
 238 CSGPRDTCFCACNPNFNSGACVTCGCPOTFYNNPTTPOLEHNPANKTYGAFYCKKPHNF 297
 304 LSTDVGSSTLVCPILHNOVTRAEQTCQKCSKPCRAVCYGLDMQYIKANSKRTITTELE 363
 298 V-VSSSSVCRAKCPSSKMEV-BENDIRKXCPCCTIDCPKACGIGTGSJMSQTVDSNIDK 355
 364 FAGCKTIFGSLAFPIESFGDGPASNTAPLPEOLOVEFTELTGVLVYSAMPDSDLS 423
 356 FINCKIKNGNLIPLVGTGHRDPYNAIIPKELNFRVYREITGELNOSWPPNPTDS 415
 424 VFQNLQVIRGILHNGAVSLTLQGLISWGLRSLRSLGSLALIHNNHTLCPVATVMD 483
 416 VFSNVLVIGGVVLSVGLSLILKQOGITSLQFOSLKEISAGNIYITDNSNLCTYHTINMT 475
 484 QLFNPHQALHTANREDECVEGALCHQLCARGMCGPFGQVNCQFLRGQCVBE 543
 476 TLFSTIMQRIYIRNKAENCTABGVCHNLCSGDCGKGPPQCSKRFSGRCLDS 535
 544 CRVIGQLPREYVNAHCLPCHPCOP-ONGSVTCFGEADQCVACAHYDPPFCVAPCS 602

```

Db      536 CNIYGEFREFENGSLCEHCDPOCEKMEGDLITGHPGPBNDNCTKSHFKDGNCEKCPD 595
Qy      603 GYKPDLSYPMKFPDEGACQPCPINTCHSCVDDIDKGC-----PAGCASPL 651
Db      596 GUGANSF--IFRYADPDRCHGCHPCNCTGNGTSHDCIYPTWGTSHLPGHAR-TPL 652
Qy      652 TETVSANV-GILLVYLVGVVFGILLKRRQOKRKTYMRLLLOETELVPLTPSGAMPNOA 710
Db      653 --IAAGVIGLFLVLVGLTFVAVYVRKSIX-KKELRRFL-ETELVPLTPSGTAPEOA 706
Qy      711 OMRLKEQYIKANSKFTIGTEL-TYKGIWIPDGENVNIPIAYKILRENTSPKXNKELD 769
Db      709 QURLKXETELK-RVAYKLGSGAGCTYKGIWVEGTVAIPIAKILNTTPPKANVEFMD 767
Qy      770 EAVNAGVSPYVSRLLGICLTSTVQVLTQMLPYCCLLDHYRENRGRSGODLLNMCQI 829
Db      768 EALIMASMDHPHVLVLLGVCLSPITQLVTLMPHGCILEYHEHKDNTGSLLLNMCQI 827
Qy      830 AKGMSYLEEDVRLVHNDLAARNVLSKPNHVKITDGLARLIDIDETEHADGKVPYTKM 889
Db      828 AKGMYLEERLVRHDLAARNVLSKPNHVKITDGLARLIDIDETEHADGKVPYTKM 887
Qy      890 ALESILRRFTQSDVWSYGVTVWELMTFGAKPYDGIPIREIPDLEKGERLPQPICTI 949
Db      888 ALECIHYRKFTQSDVWSYGVTVWELMTFGAKPYDGIPIREIPDLEKGERLPQPICTI 947
Qy      950 DYVMIMVKCMIMIDSECRPRFRFLVSEFRMADDPQRFVYIIONED-LGPASPLDSTFYRSL 1008
Db      948 DYVMIMVKCMIMIDSECRPRFRFLVSEFRMADDPQRFVYIIONED-LGPASPLDSTFYRSL 1007
Qy      1009 LEDDDMDGLDAEEVLVDPQGFPCDPAPAGAMVHHRSSSTRSGGDLTGLLEPSEE 1068
Db      1008 LEDDDMDGLDAEEVLVDPQGFPCDPAPAGAMVHHRSSSTRSGGDLTGLLEPSEE 1068
Qy      1068 EAPRS-----PLAP-SEGAGSDVDFGDLGMAKALQ 1099
Db      1056 YTPMSGNQFYVADGGFAAEQGVSVYRAPITSTIPEAPVAGATAEIFDDSCNGTLARKVP 1115
Qy      1100 SLPTDPSPLQRYSEDPTVPLPS-----ETDGVVAVLTCSPPEVYNQPDVAPQPPSP 1152
Db      1116 APHVEDBSTQRYADPTVPAPERSPEGLDEBGMTPMRDXPKQVETLNPVE----- 1167
Qy      1153 REGPLPAARPAAGATLERAKTLPFGNGVYKDVAFGAVENPEYLTPOGGAAPQHPPPA 1212
Db      1168 -ENPFVSR-----KNGDLQ-----ALDNPETHANMSG-----PPKA 1198
Qy      1213 -----FSPAFLNLYWDDPPERGA--PPST 1236
Db      1199 EDEYVNEPLYNTFANTYLGKAEYLNKNILNSPEKAKKAFDNPDMYHNSLPPRSTLOHPDY 1258
Qy      1237 FKGTPT-----AENPEYL 1249
Db      1259 LQESTKYFYKQNGRIRPIVAENEYL 1285

```

```

RA Marchionni M.A., Kelly R.A.;
RT "Neuregulin promotes survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes."
RL J. Biol. Chem. 273:10261-10269(1998).
RN [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system."
RL Neuron 6:691-704(1991).
RN [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=spinal cord;
RX MEDLINE=91784212; PubMed=9030624;
RA Carroll S.L., Miller M.W., Fritchett P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulin and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration."
RL J. Neurosci. 17:1642-1659(1997).
CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULIN, NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTAK. INTERACTS WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL: AF041838; AAC08899.1; -
CC DR EMBL: U52531; AAC53051.1; -.
CC DR HSRF: P11362; IRGK.
CC DR InterPro: IPR000494; EGFR_L domain.
CC DR InterPro: IPR000719; Euk_Pkinase.
CC DR InterPro: IPR002174; Furin-like.
CC DR InterPro: IPR001245; Tyr_Pkinase.
CC DR InterPro: IPR004019; YLP_motif.
CC DR Pfam: PF00757; Furin-like; 1.
CC DR Pfam: PF00063; Pkinase; 1.
CC DR Pfam: PF01030; Recep_L_domain; 2.
CC DR Pfam: PF02757; YLP; 2.
CC DR PRINTS: PR00109; TYRKINASE.
CC DR ProDom: PD000001; Euk_Pkinase; 1.
CC DR SMART: SMO0261; FU; 4.
CC DR SMART: SMO0261; TYKIC; 1.
CC DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
CC DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
CC DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
CC SIGNAL
CC CHAIN 1 25
CC FT 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
CC FT 26 651 EXTRACELLULAR (POTENTIAL).
CC FT 652 675 POTENTIAL.
CC FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).

```

```
FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 CYS-RICH.
FT NP BIND 718 985 PROTEIN KINASE.
FT BINDING 724 732 APP (BY SIMILARITY).
FT ACT_SITE 751 751 APP (BY SIMILARITY).
FT 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 242 BY SIMILARITY.
FT DISULFID 249 256 BY SIMILARITY.
FT DISULFID 262 269 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1188 1188 (BY SIMILARITY).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 P -> Y (IN REF. 3).
SQ SEQUENCE 1308 AA; 146957 MW; D944B096A08B41 CRC64;
```

Query Match 42.1%; Score 2868; DB 1; Length 1308;
Best Local Similarity 44.3%; Pred. No. 8, 1e-148;
Matches 597; Conservative 191; Mismatches 398; Indels 162; Gaps 30;

```
1 MELA-ALGRWGLL--ALLPQASTVCCTGTDMKRLRLPASPETHLDMRLHYQGCVQV 57
1 MKLATGLWVGSLVLAARTVQPSASQSVCACTENKLSLSLEQVYALKKYIENGEVVM 60
58 GNLELTYPTNASLFLDIOEVGYVLIANOVROVPLORLRIVRGTOLPEDNYALAVL 117
61 GNLEITSEHNHDSFLRSIREVGYVALNOPRYPLENTLRIRIIRGTQYERVALAIR 120
118 DNGDPLNNTTPTVGTASPGGLREILQLRSYTEILKGGVLIQRNPLCCYQDPTLUMDIFKHN 177
121 LNYRKDGNF-----GLOELGAKNITELINGSGYVVDQMKELCYADITHWQIVRNEM 171
178 QLALTLLIDITNRSRACHPCSPMCKSRCKGSESEDQSLRTVCAAGC-ARCKKPLPTDCC 236
172 PSNMTLVSTIGSSGGRCRCKSCTG-RCWGPTENHCQTLRTVCAEQCDGRCGYPRVSDCC 230
237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYVNTDPESMNPGRRTTFGASCY 296
231 HRECAGSGSGEKDIDCFACMNFNDGACVTCQPTFYVNPPTTQLEHNNAKTYGAFVCV 290
297 TACPYNVLTSDVSGCTIVCPILHNGEVTAEADGTORCEKSCPKARVCYGLMOQYIKANSKF 356
291 KKCPHNFV-VDSSSCVCRACPSKMEV--EENGIKWCXKCTCTDICPAACDGIQTSLMSAQTV 348
```

```
QY 357 IGTLEFAGCKKIFGSLAFLPESHSDGPASNTAPLQPEOLFETTELITGYLISAMP 416
DB 349 DSSNDKFINCKKINGLIFLTGHSVPYNAIDALDEKLVNTRVREITGFINICTPP 408
QY 417 DSLPLSVFQNLQVIRGRILHNGAVSLTLQGLISWGLRSRLSGSLAIHNTHLCF 476
DB 409 PMMTDFSVSNLVTIGRVLVSGSLILILKQGGITSLQFQSLKEISAGNIYITDNSLVCY 468
QY 477 VHTVWDOLFRRPHALLHTANRREDECVGGGLACHQLCARGHGWGGRPTQCNVCSQFLR 536
DB 469 YHTIMVTLTFSTVNRIVIRDNRAENCTAGWCNHLCSNDGCMGGRPDCLSCRSFSR 528
QY 537 GQECVEGRVQLGLPREYVNAHCLCPHPECOP-QNGSVTCFGPEAQCVCAHYKPPF 595
DB 529 GVICIESCNLDIGFEREFENSGICECDQSGEKMEDELITCHGGRPNCTKCSFKXGPN 588
QY 596 CYARPSGVKPDLSYMPJWKFPEDEGACOPCPINCTHSVCLDDKGC-----PA 644
DB 589 CEKCPDVLQGANSP--LFXADODRECHPCHPNCTOGCNGPTSHDCIYYPMTHSTLPQ 646
QY 645 EQRASPLTISAVV-GILVVVVGIVFGILIKRQOKIRKYMRLLOETELVEPLTPS 703
DB 647 HAR-TPL--IAAGVIGLFLIVMLTFAYVRKSKT-KKRALRPL-ETELVEPLTPS 701
QY 704 GAMPQAQMRILKEGYIKANSKFIGITEL-TVYKGIWIPGENYKIPVAIKVIRENTSPK 762
DB 702 GAPQAQDLRIKETELK-RVKVLGSGAFGVYGVIMVESETYKIVALKINLETTGPK 760
QY 763 ANKELIDRAYMAGVSGSYVRLIGITCTSTVOLVTQIMPYGCLLDHYRENRGLSGDL 822
DB 761 ANVEWBDALIMASVDHVLVRLGLVCLSPITQVLVTQIMHGLCLLEVHEHKNIQSLL 820
QY 823 LNMQCIKAGMSYLEDEVLRVARDLAARVLYKSPNHKITDFGLARLLD-DETEYHADG 882
DB 821 LNMQCIKAGMNYLEERLVARDLAARVLYKSPNHKITDFGLARLLDDETEYHADG 880
QY 883 KVPITMALLESILRRRFTHQSDVSYGYTWELMTFSKAPYDGIIPAREIDLLLEKGRLL 942
DB 881 KVPITMALLESILRRRFTHQSDVSYGYTWELMTFSKAPYDGIIPAREIDLLLEKGRLL 940
QY 943 QPCTIDIVYIMVCMWIDSECPREELVSESRMARDPORFVYVQND-IGPASPLD 1001
DB 941 QPCTIDIVYIMVCMWIDSECPREELVSESRMARDPORFVYVQND-IGPASPLD 1000
QY 1002 STFYKSLLEDDMDGLVDAEYLVPOQGFCCPP-----APG 1038
DB 1001 SKFQNLLEDEDDMDDAEYLVV-QAFNIPRTYTSRTRIDSNREIGHSPRPATPM 1059
QY 1039 AGMWHNHRSSSTRSGGDLTLGLSESEEARPSPLAPEGAGSDVFDGLMGAAKGL 1098
DB 1060 SSGQFVYODGFATQOG--MPMPTATSTIPAPVA--QGAIAEMFDDSCCNGLRKP 1114
QY 1099 QSLPTHPSPLORESDEPTVLPFS-----ETDGYAPLPTCSPOPEYVVPVRPQPS 1151
DB 1115 VVPHVQEDSSQIRASADPTVFAPEARNRALLDEGGYMTPHMDPKQGEYLNPE----- 1167
QY 1152 PREGRPLPARAGATLEBAKTLPSGKGVKVDVAFAGCAVENPEYLTPOGGAAPQHPRP 1211
DB 1168 --ENPFVSR-----KNGDIQ-----ALDNPEYSASG-----PK 1197
QY 1212 A-----FSPARDNLYYMDQDPERGA--PPS 1235
DB 1198 AEDRYVNEPLVLTFTNALGNAEYMKRSLSVBEKAKKAFDNDPMHSLPRPSTLQHPD 1257
QY 1236 TFKGPT-----AENPEYL 1249
DB 1258 YLOEYSTKYFYKQNGRIRPIVAENPEYL 1285
```

RESULT 8
XMRK_XIPMA
ID XMRK_XIPMA
AC P13388;
STANDARD; PRT; 1167 AA.

FT	DISULFID	237	245	BY SIMILARITY.
FT	DISULFID	241	253	BY SIMILARITY.
FT	DISULFID	256	265	BY SIMILARITY.
FT	DISULFID	269	286	BY SIMILARITY.
FT	DISULFID	300	311	BY SIMILARITY.
FT	DISULFID	315	330	BY SIMILARITY.
FT	DISULFID	333	337	BY SIMILARITY.
FT	DISULFID	504	513	BY SIMILARITY.
FT	DISULFID	508	521	BY SIMILARITY.
FT	DISULFID	524	533	BY SIMILARITY.
FT	DISULFID	537	553	BY SIMILARITY.
FT	DISULFID	566	569	BY SIMILARITY.
FT	DISULFID	560	577	BY SIMILARITY.
FT	DISULFID	633	615	BY SIMILARITY.
FT	DISULFID	618	626	BY SIMILARITY.
FT	CARBOHYD	622	634	BY SIMILARITY.
FT	CARBOHYD	114	114	N-LINKED (GLCNAC .) (POTENTIAL).
FT	CARBOHYD	144	144	N-LINKED (GLCNAC .) (POTENTIAL).
FT	CARBOHYD	201	201	N-LINKED (GLCNAC .) (POTENTIAL).
FT	CARBOHYD	356	356	N-LINKED (GLCNAC .) (POTENTIAL).
FT	CARBOHYD	365	365	N-LINKED (GLCNAC .) (POTENTIAL).
FT	CARBOHYD	398	398	N-LINKED (GLCNAC .) (POTENTIAL).
FT	CARBOHYD	417	417	N-LINKED (GLCNAC .) (POTENTIAL).
FT	CARBOHYD	501	501	N-LINKED (GLCNAC .) (POTENTIAL).
FT	CARBOHYD	576	576	N-LINKED (GLCNAC .) (POTENTIAL).
FT	CARBOHYD	621	621	N-LINKED (GLCNAC .) (POTENTIAL).
SQ	SEQUENCE	1167 AA;	129934 WM; 4793E0749DC1D55A CRC64;	
Query Match 38.4%; Score 2614.5; DB 1; Length 1167;				
Best Local Similarity 44.4%; Pred. No. 3.7e-134;				
Matches 564; Conservative 167; Mismatches 397; Indels 141; Gaps 31;				
OY	AALCRGSLALLPRGAASST--- <td>59</td> <td></td> <td></td>	59		
DB	8 AALIQ--LLVLVISIRCCSTDPRDKRCQGCTSNMQTM--LDNHYLKKMKXMYSGCNAVLEN	62		
OY	LELTATPLPNASTLSDIOIEVGQGVLIANHQVRPEQRRIIVGTQLFFENYLAVALVDN	119		
DB	63 LEITTOQNOLDLSLOSIQEVGGVVLIAMNEVESTIPLVNLRLRGONLYEENFLTYMSN	122		
OY	120 GDPINNTPTVYGASPGGLRELOLSTELIKGYLIQRNPOLCYDTIIMWDIPHKNQL	179		
DB	123 YOK-NPSSP--DYQVGLKGQLQSNI LLEILLISGGVKYSHNPLLNCVEITIIIMWVIDKTSP	179		
OY	180 ALTILDTRKSAAICPCSEPMCKGSFCMESESBDOSLTRVCAAGC-ARKGPPLPTDCHE	238		
DB	180 TMLNLIPIAFERQCKCHDHGCNVGSCMAPRPHQCKCFKLTLCAEQCNNRCGRPKRIDCNE	239		
OY	239 QCAAQCGTPKSIDCLAIHFNSHCIGLHPALVTYTDPFEPSMPEGRTTGASCUTA	296		
DB	240 HCAGGCTGFPAITDCLAIRDNDSDTCDDTPRPKIYDIYHQVVDNNINRYTTGAACVKE	299		
OY	299 CPVAVYSTDVSGCTLVCPRLNQEVTAEDGTORCEKSKPCARVICYGLGMQVIKANSKFIC	358		
DB	300 CPSNYVYTE-GACYRSCSAGMLEVD-ENGKRSCRKPCDGVPKYCDGDIGISTL-SNTIAVN	356		
OY	359 ITTEL-EFAAGCKIFGSLAPLESDDGPASANTAPLQEOUQEETILEITGIYLIAMPD	417		
DB	357 STNRSSNSTCKINGDIILNRNSEBGPXYICTMPBEHLMTLTKYEITGIYLIWMWE	416		
OY	418 SLPLDSVFONIQUIRGRILHNNGAYS--TLIOGLGISWLGLSRLEISGLAILHHNTHLCF	476		
DB	417 NMTELSYFOUILEIRGHTTSRGHSFVVGVRLQWLGRLSLVEVASGANVILIKNTIQLERY	476		
OY	477 VHVTPMWQLFRNFHQALIHATANPEDECSVEGLAQHLCARGCMGRRPTOCVNGSQFLR	536		
DB	477 ANITNMRLRSSEDQISEYDART-----ENDTCNNCESDEDSCWBPGMTCVSCIAHVR	529		
OY	537 GOECVEBCRWLQGLEPREYVARJCLPCHEPCQONGSVYICFGHEADQCAVAAYXPDPFC	596		
DB	530 GRCVNASCNLLQGEPRFAQYDGRCSVCHOECVLQTDSDSLTCYEGGRANCKSAHFQGGPOC	589		
OY	587 VAARDREIKENTSYMITWRPFEEFYACGCCGAGVIGTCTGAEAGCATATGATTATGAC	589		


```
FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.
FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 266 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 552 565 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 141 183 EILSGVYIEKNDLCHMDITDMEDIYRDADAEIVKNGR
SC -> GGFVWVSGLTPOADMDYLLDDPRLLTTSASX
VPVTLAAV (IN SHORT ISOFORM).
FT VARSPLIC 184 1342 MISSING (IN SHORT ISOFORM).
FT CONFLIT 560 560 E -> G (IN REF. 2).
FT CONFLIT 1064 1064 E -> G (IN REF. 2).
SQ SEQUENCE 1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;
```

Query Match 34.4%; Score 2341.5; DB 1; Length 1342;
Best Local Similarity 39.6%; Pred. No. 2.5e-119;
Matches 521; Conservative 197; Mismatches 460; Indels 137; Gaps 36;

```
OY 10 GLLALLPPGAA--STOYCTGTDMKRLPASPTHLDMLHLYOGQVQVGNILETYLPT 67
DB 11 GLFLSARGSEVNSGAVCPGTLINGLSVTGDAENQYQTLKXIVERCEVWGNILEVLTGH 70
OY 68 NASLSELODIENOGVYLIANNOVROVPLORLRVYRGTOULFEDNYALALDNGDPYLNNT 127
DB 71 NADLSFLQWIREVTVLVANNEPSTLPLNLRVYRGTOYDCKFAIFWV----LNYNT 125
OY 128 PVTGASPGGLRELOLRSLITELIKGVLIOKRNPOLCYODITLWKDIFHKNNQALATLIDTN 187
DB 126 ---NSSHALRQLLTQLTTELISGVYIEKNDKLCMDITDMEDIYRDRD---AEIVVD 178
OY 188 RSRACHPCSPWCKSRCKWGESSECCSLTITVACGGC-AACKPPLPTDCCHEGCAAGCTG 246
DB 179 NGRSCPCHVEVCKK-RCKWPGSEDCQTLTKTICAPQCNCHGCFSPNNQCHCEBACAGCGSG 237
OY 247 PKHSDDLCLAHFNHSGICEJHCPALVTYNTDFESMNPGRGTFGASCVTACFVNYLST 306
DB 238 PQDTCDFACHFNHSGACVRCRCPQVLYNKLITQLEBNPTTKIQYGVCAVACPHNFV-V 296
OY 307 DVSGCTLVCLPILHNOEVTADGTOECCKSPKCAVCGYGLQMOYIKANSKP--IGITELR- 363
DB 297 DQTSQVRACPEDKKEVD-KNGLKWCCEPCGGLCPKACEGTV-----SGSRFQTVDSGNDIG 350
OY 364 FAGCKTIFGSLAFPESEFDGDPASNTAPLOPELOVPELTLEETGLYIYSAMPDLSPLDS 423
DB 351 FVNCSTKLGNLDFLITLNGDPWPKHKIPALDPEKLNFRVTRISLTGLNLQSWPHEHNS 410
OY 424 VPQNLQVIRGRIIHNAGVS-LTQGLGISWLGRLSRLREGSGALIIHNTHPCFYVATPW 482
DB 411 VFSNLTITIGRSYINRGFSILIMKNLIVTSLGFRSLKEISAGRIYISANROLCYHHSLSNW 470
```

```
OY 483 DQLEFNPHQALLHTA-NRPEDECVGEGLAQGLCARGCHWGPGPTQCVNCSQFLRGBCV 541
DB 471 TKVLGPTFEERLDIKANPRRDCVABGVCCPLCSGAGCGPGGQLSKRNSRGVGV 530
OY 542 EECRVLOGLPREYVNAHRLCPCHRECOPONGSVTCFPGPADOCVACAYKDPPEVCARCP 601
DB 531 THCNPLNGEPREFAHEACFSCHRECOPMEGTATCNSGSDTCAQCAHPDGPCHVSCSP 590
OY 602 SGVRKDLSTYMIKMFPEBEAGCQCPINCHTSCVDLDDKCPAEQRA-----SPLTSTVSA 657
DB 591 HGVLG--AKGTYKPYDQNECRCHENCTOGCGPBLQDCLGTVLIGKTHLTMTALV 648
OY 658 VGLILVVLVGVVEGILIKRQOKIR-KYTRRLQETLEVEPLTPSGAMPNOQWMLK 716
DB 649 IAG--LVVIFMMLGGTFLYMGRIQKRAMRLTEGSETEPLDPS-EKANKYLARIK 705
OY 717 EOYIANSKFIQITEL-TVYKGIWIPGENVKIPVAIKVIRENTSPRANKEIDEAYVMA 775
DB 706 ETELR-KLKVIGSGVFGVHKGVMIPEGESIKIPVCIKVIEDKSGROSFOAVTDHMLAIG 764
OY 776 GVGSPYVRRLIGICTSTVOLVTQLMFYGLLDHYENRGRIGSODLNNCMQIAKMSY 835
DB 765 SLDNAHIVRLGICPGSSLOLVTOYLPGLSLDHRQHGALGQLLNMGVOIAKMY 824
OY 836 LEDVRLVHRDLAARNVLVKSFNHYKIDFGLARLJDDETEYHADGSKVPDKMALESIL 895
DB 825 LEEHGMVHRNLAAARNVLKSPQYQVADLFPDQQLLYSEKKTIKMALESIH 884
OY 896 RRRPETHODWVSYGVTWELMTFAKYYDGIPAETFDLLEKGBRLFOPRTCTIDYVMIM 955
DB 885 FGKTYHSDWVSYGVTWELMTFAEYVAGLRLAEEVDLLEKEBERLAQPCICTIDYVMIM 944
OY 956 VKCMWIDSECRPRELIVSEFSERAPDPORFVYIOWEDLGRA---SPLDSTFYSLLEDD 1012
DB 945 VKCMWIDENIRPTKELANETFRARPPRYIVIKES-GPGLAPGBEPGLNKKLIEV 1003
OY 1013 DMGDLVDAEEVLYQQGFPCPDPAAGVHHHRHSSSTRSGGGLTGLAP-SEEAR 1071
DB 1004 ELEBEDLDLDLEAED-----NLATTIGSALSLPVGTLNPRGS 1044
OY 1072 RSPFAPSEGAGSDVFPDDDLGMAKGLQSLPTH-DPSLQRYSEDDPTVLP-----SET 1124
DB 1045 QSLSPSSGY-MPMNQNLBESCOESAVSSSERCPPVSLR-----PMPRGCLASSESS 1097
OY 1125 DGYYA-----PLTCSPOPE-----YVNPQVRRPQPSPREGP----- 1156
DB 1098 EGHVTGSEAELOEKVSMCRSRSRSPRPGDGAHNSQHSLLTPTPLSPGLEBEDVN 1157
OY 1157 ---LPARPAGATLEAKTISP-GKXGV-----KDVFAFGAVENPEYLTQGGAAPO 1206
DB 1158 GYVMPDTHLKGTPSSREGTLSSVGLSLGTBEEDD-----EEYEYMNRRRHSP- 1208
OY 1207 PHPPEAFSPAFDNLVYWD-----QDPPEGAPSTFGTPTAENPEYL 1249
DB 1209 PHPPEPSLDELGEYNDVGSDSLASLSTQSCPLPVPIMPTAGITTPBEDVBYM 1263
```

RESULT 10
ERB3_RAT STANDARD; PRT; 1339 AA.
AC Q62799; O62955;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (C-erbB3).
GN ERB3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.

CC STRAIN-Sprague-Dawley; TISSUE=Liver;
 RX MEDLINE=96086535; PubMed=8522190;
 RA Hellyer N.U., Kim H.-H., Greaves C.H., Sterke S.L., Koland J.G.;
 RT "Cloning of the rat ErbB3 CDNA and characterization of the
 RT recombinant protein.";
 RL Gene 165:279-284(1995).
 RN (2)
 RP REVISIONS TO 85: 513 AND 565.
 RA Hellyer N.U., Koland J.G.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE OF 922-1097 FROM N.A.
 RC STRAIN-Sprague-Dawley; TISSUE=Sciatic nerve;
 RX MEDLINE=97184212; PubMed=9030624;
 RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
 RT "Expression of neuroligin and their putative receptors, ErbB2 and
 RT ErbB3, is induced during Wallerian degeneration.";
 RL J. Neurosci. 17:1642-1659(1997).
 CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
 CC AND PROMOTES ITS ASSOCIATION WITH THE P65 SUBUNIT OF
 CC PHOSPHATIDYLINOSITOL 3-KINASE.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U29339; AAC28498.2; -
 DR EMBL: U52530; AAC33050.1; -
 DR HSSP: P11362; 1PKK.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR01245; Tyr_pkinase.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR PRINTS: PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; Fuf_5.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS0011; PROTEIN_KINASE_DOM; 1.
 KM Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 19
 FT CHAIN 20 1339
 FT DOMAIN 20 643
 FT TRANSMEM 644 662
 FT DOMAIN 663 1339
 FT DOMAIN 183 259
 FT DOMAIN 707 964
 FT NP_BIND 713 721
 FT BINDING 740 740
 FT ACT_SITE 832 832
 FT DISULFID 186 194
 FT DISULFID 190 202
 FT DISULFID 210 218
 FT DISULFID 214 226
 FT DISULFID 227 235

FT DISULFID 221 243 BY SIMILARITY.
 FT DISULFID 246 255 BY SIMILARITY.
 FT DISULFID 259 286 BY SIMILARITY.
 FT DISULFID 290 301 BY SIMILARITY.
 FT DISULFID 305 320 BY SIMILARITY.
 FT DISULFID 323 327 BY SIMILARITY.
 FT DISULFID 500 509 BY SIMILARITY.
 FT DISULFID 504 517 BY SIMILARITY.
 FT DISULFID 520 529 BY SIMILARITY.
 FT DISULFID 533 549 BY SIMILARITY.
 FT DISULFID 556 573 BY SIMILARITY.
 FT DISULFID 576 585 BY SIMILARITY.
 FT DISULFID 589 610 BY SIMILARITY.
 FT DISULFID 613 621 BY SIMILARITY.
 FT DISULFID 617 629 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 459 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 1028 1028 L -> P (IN REF. 3).
 SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F4028BDFE1E CRC64;
 Query Match 33.3%; Score 2269.5; DB 1; Length 1339;
 Best Local Similarity 40.0%; Pred. No. 2e-115;
 Matches 514; Conservative 173; Mismatches 438; Indels 161; Gaps 37;
 3 LAALRWGGLLALPPGAA---STQVCGTGMKRLPASPETHLDMRHLYOGGVQVGN 59
 7 LQVLC-----FLLSLARGSEMNSQAVCGTLNGLSTVDGADNVOYQTLKYKECEVMA 62
 60 LELTVLPFNASLSFQDIQEVGYVLLAHNQVRQVPLQRLRYRGTLFEDNYALAVLDN 119
 63 LEIVLTGHNADLSFLQWIREVTGYLVANNEFSLPLNLRVVRGTQVQGFALFWM-- 120
 120 GDLPLNNTFPVVGASPGGIRELQSLSTREILKGGVLLIQRNQLCYQDTILMKDIFHK 179
 121 ---LNYNT-----NSSHLRQKLPQQLTEILSGVYIKENDLCMDITDRDLYVR-- 170
 180 ALTLIDNRSPACHPCSPMCKSGRSCWSESSDCSLTRTVACAGC-ARCGPLPTDCHE 238
 171 GAELVKNKNGANCPCHEVCKG-RQWGGPDDCOILKLTICAPQCNCRCGPNQCCHD 229
 239 QCAAGCTGPKXSDCLACLFPHNSGICELHCPALVYNTDTFESMPNDEGRYTGASC 238
 230 ECAGGCGSPDDTDFACRFRFNDGACVPRCEPLVYVKLTFOLEPNPHTRYQVGVVAS 289
 299 CPVNYLSTDVSGCTLVCPLNQETAEQDRCCEKCPARVCYGL--GMQYTKANSKF 356
 290 CPHNV-VDOIFCYRACPPDKMEVD-KHGLKCEPCGGLCPKACEGSSSRQYTDSSN 347
 357 IGTLEPAGCKKIPGSLAFIPESFDQPSANTAPLOPELOVFEETLEITGYLYISAWP 416
 348 ID---GFVNCTKILGNIDFLITLNDVPMHKIPALDPEKLVNFRVREITGYLNIQSWP 403
 417 DSLPLSTFQNLQVIRGILHNGAYS-LTLQAGISMLGSRISREISGSLALHNHTHLC 475
 404 PHMNFHSFSLUTTIIGRSLYNRFSLLIMKANLVSTLGRSLKEISAGVYISANQQLC 463
 476 FVHTVPMQDLFRNPHQALLHTA-NRPDECEVGGGLACHQLCARGHWCMPGPTQVNC 534
 464 YHSHLNTWRLRLRGSEBERLDIKYDRPLGEICLAEGKVDPLCSSGGCGPDPGCLCRNY 523
 535 LRQGEVCEECVILQGLREYVNAHCLCPHPEQCPQNGSTTCGPREDDQVAAHAKDP 594
 524 SREGVCTVHCNFFLQGBREFVHAQCFSCHPECLPMEGISTCGSSSDACARCAHFRDGP 583
 595 FCVARCBSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSC--VDLDKCGCAPQASPLT 652

```

DB      584 HCVNSCFPIIG--AKGPITYKYPDAONECEPCHEHCNTOGCNGPELQJCLGQAEVLMSKPH 641
QY      653 SIVSAVVGILLVVLGVAVFGLIRKROQKIR-KYMERLLQETVELVPLPSSAMPRQAO 711
DB      642 LVIANTVG--LAVILMTIGSFLYWRGRIONKKAMRRYIERGESIEPLDPS-ERANKVL 698
QY      712 MRILKEQYIKANSKFEIGTEL-TYKGIWIPDGNVYPAVAKYLRENTSPPKNEKILDE 770
DB      699 ARIFKETELR-KLKVAGSGVGFVHKGIWIPEGESIKIPICIKYIEKSGRCGFQAVTTH 757
QY      771 AYVAVGSPVVSRLIGLCITSTVQLTQCLMPVGCILDHYRENNRGRIGSDLLNMCQIA 830
DB      758 MLAGSLDHAHIVRLGLCGPSSSLQVLTQVLPGLSLDHYKQHEBTGLPOLLNMGVQIA 817
QY      831 KGMSTLEDVRLVHDLAARNVLRKSPNHVKTTPDGLARLIDIBETEXHADGKVPYKMA 890
DB      818 KGMYYLEHSHVHNDLARNVLRKSPNHVKTTPDGLARLIDIBETEXHADGKVPYKMA 877
QY      891 LESILRRRFTHQSDVSVYGVTVWELMTFGAKPYDGIARSIPLLEKGERLPQPICTID 950
DB      878 LESIHFGKYTHQSDVSVYGVTVWELMTFGAEYAGRLAEIPDLEKGERLAQPICTID 937
QY      951 VYMIWVKCMIIDSECPREFELVSEFSRMAPDQRFVYIQNEIDGPAISPLDSTFFYSLE 1010
DB      938 VYMWVKCMIDENIRPFFELANEFTRMADPFRVYVIRAS-GPQTP--PAAPESVLT 994
QY      1011 DDMGDVLDAEEYLVPOQGFPCPDPAAGAGVYHRRSSSTRSGGDLTLGLEPSEE-- 1068
DB      995 TKEL-----GEAELEPEL-----DLDLDLAEHEGL 1020
QY      1069 -----EAPRSPLASEG-----AGSDVFDGDLGMAKKGITQSLP 1102
DB      1021 ATSLGASLSLTGTLTTPRGQSLSLSPSSGYMPMNGSLDEACLDASVYAGRGREFSRPS 1080
QY      1103 THDPSPLQRYSEDPVTLPLSETDGVY-----APL-----TC-----SPOPE---YVNG 1142
DB      1081 LH-FIPRGR-----PASSESECHVYSGAELEQEKVYSRSRSRSRSPRFRDSAYHQ 1132
QY      1143 PDVVRPQPPSPREGP-----LPAPPAQATLERAKTISP-GKNGVY-----XDV 1184
DB      1133 RNSLTLPTTPSPFGLEEDDNGVYMDTLRGASSREGTILSVGLSVLTGHEEDED- 1191
QY      1185 FAFGAVENPEYLLPQGAAPQPPR 1210
DB      1192 -----EVEHYMKRKRGRSP-PRPP 1209

```

```

RESULT 11
EGFR_DROME
ID EGFR_DROME STANDARD; PRT; 1426 AA.
AC P04413; O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DB Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DB (Gurken receptor) (Torpodo protein) (Drosophila relative of ERBB).
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryote, Metazoa, Arthropoda, Mandibulata, Pancrustacea, Hexapoda;
OC Insecta, Pterygota, Neoptera, Endopterygota, Diptera; Brachyera;
OC Muscomorpha, Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN 11
RP SEQUENCE FROM N.A. (ISOFORM TYPES I AND II).
RX MEDLINE=84350209; PubMed=8070664;
RA Clifford R., Schubach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in
RT subdomains of the receptor protein.";
RL Genetics 137:531-550(1994).
RN 12
RP REVISIONS.

```

```

RA Clifford R., Schubach T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN 13
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlesinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains.";
RL Cell 40:599-607(1985).
RN 14
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC STRAIN=Oregon-R; TISSUE=Embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schaefer E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcripts.";
RL Cell 46:1091-1101(1986).
RN 15
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RT specification in wild-type, Ellipse, and null mutant Drosophila.";
RL Dev. Biol. 205:129-144(1999).
RN 16
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RX STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731122;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer V., Champs O., Pfeiffer B.D.,
RA Abail J.F., Agbayani A., An H.-U., Andrews-Flannoch C., Baldwin D.,
RA Baller R.M., Baub A.A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhattacharya D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brostein P., Brotier P.,
RA Butts J.K., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Chert J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dup J.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeller C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Huck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalaal M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimbel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Murry D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclev J.M.,
RA Palazzolo M., Peltan G.S., Pan S., Pollard J., Puti V., Reese M.G.,
RA Reinert C., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gkbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN 17
RP SEQUENCE OF 959-1078 FROM N.A.
RX STRAIN=Daekwanryeong;
RX MEDLINE=85137938; PubMed=2983222;
RA Wadsworth S.C., Vincent W.S.II, Bilodeau-Wenworth D.;
RT "A Drosophila genomic sequence with homology to human epidermal

```



```
DB 297 PDKMALESILHRITTHOSDVMSYGVATWELMTOSKRYDIPASEISSVLEKGERLPDP 356
QY 945 PICTIDVIMVWKCMWIDSECRPRELVSEFSRMADPQRFVVIQ-NEDLGAPSPDST 1003
DB 357 PICTIDVIMVWKCMWIDADSRPFRELIASFSCXKARDPRRYLVIOQDERMHLPSPTDSK 416
QY 1004 FYRSLLEDMDGDLVDAAEYVLPQGGFCPCDPAPGAGCMVHHRSSSTSGGSDLTLGL 1063
DB 417 FYRTLMEEDEMDIYDADBYLVPHGFF-----NSPST----- 449
QY 1064 EPSEEARPSPL-----APSEGAGSDVFDGDLGAKGLQSLPTHDPSPQRYSEDPV 1118
DB 450 -----SRTPLSSLSLATSNNSATNCID-----RNGQHPRVEDSFVQRYSDPTG 494
QY 1119 PLPSET--DGVVAPLTSCPOPEYVNPDPVPPSPRPGPLPARAPAGATLERAKTISPG 1176
DB 495 NFLESIDDGFL-----PAPRYVQ--LMPKPS-----TAMV 525
QY 1177 KNGVVKDF-----AFGAVENPEYLPQGAAPQHPAPFAPFADNLY 1221
DB 526 QNGIYNNISLTAISKLPMSDRYQNSHSTAVDNPEYL-----NTNQSPLAKTVFESSP 577
QY 1222 YWDO-----DPE-----RGAPSTKGPPTAENPEYLGDPV 1254
DB 578 YVIOGNGHQLNDNPQYQDPLNETKPNGLKVPAAENPEYLVAP 625

RESULT 13
ERBB AVIER STANDARD; PRT: 604 AA.
ID ERBB AVIER
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxId=79685;
RN (1)
RC SEQUENCE FROM N.A.
RA MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
RT gene family.";
RL Cell 35:71-78(1983).
RN (2)
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RX Debierre B., Henry C., Benaisa M., Biserte G., Claverie J.-M.,
RA Saulie S., Martin P., Stehelin D.;
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
RT new type of oncogene.";
RL Science 224:1456-1459(1984).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC / tyrosine phosphate.
CC -1- DISEASE: The V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -1- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
DR EMBL: K02006; AAA42394.1; ALT_INIT.
DR EMBL: K01216; AAA42400.1; -.
DR PIR: A00644; TVI0H.
DR HSSP: F1362; IFGK.
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR001245; Tyr_Pkinase.
DR Pfam: PF00069; Pkinase_1.
DR ProDom: PPO00001; Euk_Pkinase_1.
DR SMART: SMO0219; TykC; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76BCDD06745D609 CRC64;

Query Match 23.8%; Score 1623; DB 1; Length 604;
Best Local Similarity 50.7%; Pred. No. 7.5e-81;
Matches 350; Conservative 76; Mismatches 137; Indels 128; Gaps 18;

QY 587 CAHYKDPFCVARGCPGKPDLSYPMKPFDEGACQPCPINCSTHSCVLDKGCFAEQ 646
DB 3 CAHFDGPHCVKACPGVLGENDTL-VRKADANAVQCLQHPCTCKGPGEGCP-- 58
QY 647 RASPLTISVNAV-GILLVVLGVVFGILIKRQOKIRKYTMRLLOETLVEPLTPSGA 705
DB 59 NGSKTPSIAGVVGGLCLVYVVGIGILYRR-HYRKRTLRLLQERLVEPLTPSGE 117
QY 706 MPNQAKRLKEQYIKANSKFIGTEL-TYKGIWIPDGNVKI PVAIKYLRNTSPKAN 764
DB 118 APNQAHLILKETBFK-KVYKLGAGAGGTIYKGIWIPDGNVKI PVAIKYLRNTSPKAN 176
QY 765 KEILDEAVYMGVSPVYRLLGCLSTYQOLVQTMPYCCLDHYRENGRGOSDOLN 824
DB 177 KEILDEAVYMGVSPVYRLLGCLSTYQOLVQTMPYCCLDHYRENGRGOSDOLN 236
QY 825 WCMQJAKMSYLEDLVLRVHRDLAARNLVYS PNHVKITDGLARLLDIDETEHADGKV 864
DB 237 WCMQJAKMSYLEDLVLRVHRDLAARNLVYTPHVKITDGLARLLDIDETEHADGKV 296
QY 885 PDKMALESILRRRFTTHOSDVMSYGVATWELMTFGAKPYGIPAREIPDLLEGGERLPQ 944
DB 297 PDKMALESILHRITTHOSDVMSYGVATWELMTFGSKPYGIPAREIPDLLEGGERLPQ 356
QY 945 PICTIDVIMVWKCMWIDSECRPRELVSEFSRMADPQRFVVIQ-NEDLGAPSPDST 1003
DB 357 PICTIDVIMVWKCMWIDADSRPFRELIASFSCXKARDPRRYLVIOQDERMHLPSPTDSK 416
QY 1004 FYRSLLEDMDGDLVDAAEYVLPQGGFCPCDPAPGAGCMVHHRSSSTSGGSDLTLGL 1063
DB 417 FYRTLMEEDEMDIYDADBYLVPHGFF-----NSPST----- 449
QY 1064 EPSEEARPSPL-----APSEGAGSDVFDGDLGAKGLQSLPTHDPSPQRYSEDPV 1118
DB 450 -----SRTPLSSLSLATSNNSATNCID-----RNGQHPRVEDSFVQRYSDPTG 494
QY 1119 PLPSET--DGVVAPLTSCPOPEYVNPDPVPPSPRPGPLPARAPAGATLERAKTISPG 1176
DB 495 NFLESIDDGFL-----PAPRYVQ--LMPKPS-----TAMV 524
QY 1177 KNGVVKDF-----AFGAVENPEYLPQGAAPQHPAPFAPFADNLY 1217
DB 526 YVIOGNGHQLNDNPQYQDPLNETKPNGLKVPAAENPEYLVAP 573
QY 1218 DNLTYWQDPEPERGAPSTKGPPTAENPEY 1248
```



```
Db 601 CVKACPAGVLGENDTL-VWKYADANAVQQLCHPNCCTRGCKGPGLEGCP---NGSKTPSIA 656
QY 656 SAVV-GILLVVVLGVFGLILIKRQOKIRKTYMRLLOETELVEPLTP 702
Db 657 AGVVGGLCLLVVVGIGLGLYLR--HIVKRTLRLLQERELVEPLTP 703

RESULT 15
ERBB_AVIEU STANDARD; PRT; 540 AA.
ID ERBB_AVIEU
AC P11273;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ts167).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxId=103898;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=87064458; PubMed=2878364;
RA Choi O.R., Trainor C., Graf T., Beng H., Engel J.D.;
RT "A single amino acid substitution in v-erbB confers a thermolabile
RT phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT cells."
RL Mol. Cell. Biol. 6:1751-1759(1986).
CC -1. CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1. SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M13179; AAA42401.1; -.
DR PIR; A25231; TVFEB.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00069; Pkinase; 1.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SMO0219; TYRC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).
SQ SEQUENCE 540 AA; 60412 MW; 5B53297AA068B65D CRC64;

Query Match 22.9%; Score 1558; DB 1; Length 540;
Best Local Similarity 53.5%; Pred. No. 2.2e-77;
Matches 332; Conservative 69; Mismatches 127; Indels 92; Gaps 16;

QY 587 CAHYKDPFCVACPSGVKPLSYMPIKFPDEGACOPCPINCTHSCVDLDDKGPAG 646
Db 3 CAHFDGPHCVKACPSAGVLGENDTL-VWKYADANAVQQLCHPNCCTRGCKGPGLEGCP---58
QY 647 RASPLTSIVSAVV-GILLVVVLGVFGLILIKRQOKIRKTYMRLLOETELVEPLTPSGA 705
Db 59 NGSKTPSIAAGVVGGLCLLVVVGIGLGLYLR--HIVKRTLRLLQERELVEPLTPSG 117
QY 706 MPNQMRLIKEQYIKANSKFIGITEL-TVYKGIWIPDGENVKIPVIXLRENTSPKAN 764
```

```
Db 118 APQAHRLILKETEFK-KYKVLGPGAGTGYKGLWIMEGKVTIIPVAIKELREATSPKAN 176
QY 765 KEILDEAYVWAGVSPVYSLGLGICLTSTYQVLTQMPYGCILLDHNRENRGRIGSDLLN 824
Db 177 KEILDEAYVWASVDNPVCRLLGLICLTSTYQVLTQMPYGCILLDYNREKDNIGSOYLLN 236
QY 825 WCMQIAKMSYLEDVRLVHRDLAARNVLYSPNNVITDEGLARLLDIDETEHYAGCKY 884
Db 237 WCVQIAKQMYLERHMYHVDLAAARNVLYVTPOHVKITDGLAKQGLADEKEHYAGCKY 296
QY 885 PIKMALESILRRFTHQSDVWSYGVTVMEMLTFGAKPYDIPAREIPDLLEKGERLPQP 944
Db 297 PIKMALESILHRYTHQSDVWSYGVTVMEMLTFGSKPYDIPASEISSVLEKGERLPQP 356
QY 945 PICTIDVYMIWKKWMSDASCRRPREFVSEFARMADPQRFVVIC-NEDLGSPASPLDST 1003
Db 357 PICTIDVYMIWKKWMSDADSRPFRELIASFKMADDPRIYLVIGDERMHLPSFTDSK 416
QY 1004 FYRSULEDDMDGLVDAEYLVDPQGFPCPDPAFGAGVYHRRSSSTRSGGDLTLGL 1063
Db 417 FYRTLMEBEDMEDIVDADEVLPHQGF-----NSPST-----449
QY 1064 EPSEEARPSFL-----APSEGASDVFDGDLGMAKGIQSLPTHPDPSFLQYSEDPVY 1118
Db 450 -----SRTPLLSLSLATSNNSATNCIDRNG-----H-----476
QY 1119 PLPSETDGYVAPLTCSPQPEYVQNPVRPOPSPRECEPLPAAPAGAT-LEBAKTLSPCK 1177
Db 477 --PYREDOFL-----PAPRYVNG--LMPKKPSTAMQNGIYVYISLTASKLPIDSRVQ 526
QY 1178 NGVVKDVPAFGGAVENPEYL 1197
Db 527 N-----SHSTAVDNPEYL 539
```

Search completed: July 22, 2003, 09:18:31
Job time : 20.642 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:01:42 ; Search time 46.8246 Seconds

(without alignments)
5522.503 Million cell updates/sec

Title: SEQ4-695-709-12

Perfect score: 6815

Sequence: 1 MELALCRWGLLALPFGA.....TFKGTPTAENPEYLGDVY 1255

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: SP_archaea:*
2: SP_bacteria:*
3: SP_fungi:*
4: SP_human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phase:*
10: SP_plant:*
11: SP_prodent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*
15: SP_rvirus:*
16: SP_bacteriap:*
17: SP_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6129	89.9	1259	6	018735
2	3065	45.0	1209	11	090X70
3	3036	44.5	1210	11	09EP98
4	2656	39.0	1165	13	09YH40
5	2615.5	38.4	1137	13	09W6F6
6	2225	32.6	1328	13	P79754
7	1971.5	28.9	1433	5	09B1H9
8	1871	27.5	419	4	09UK79
9	1739	25.5	367	11	08R2X1
10	1697.5	24.9	412	4	08WV70
11	1640	24.1	729	15	08E712
12	1638	24.0	567	15	08E712
13	1581.5	23.2	567	15	08E712
14	1573	23.1	545	15	08E4895
15	1506.5	22.1	655	11	09WVFS
16	1490.5	21.9	643	11	09ERV6

17	1211	17.8	1193	5	09Y1X8	09Y1X8 ephydactia f
18	1146.5	16.8	1368	5	023821	023821 caenorchabdi
19	1134	16.6	1717	5	026566	026566 schistosoma
20	1126	16.5	527	13	090836	090836 gallus gall
21	1001.5	14.7	478	11	09E8E0	09E8E0 gallus gall
22	942.5	13.8	599	13	09E8E2	09E8E2 gallus gall
23	906	13.3	165	4	014256	014256 homo sapien
24	887	13.0	176	11	0923V5	0923V5 ractus norv
25	806.5	11.8	346	13	P11776	P11776 xiphophorus
26	778	11.4	435	5	08SZM1	08SZM1 drosophila
27	754.5	11.1	311	13	099162	099162 xiphophorus
28	734	10.8	331	4	09BD77	09BD77 homo sapien
29	723	10.6	149	6	09B366	09B366 oryctolagus
30	706.5	10.4	1362	11	09PVZ4	09PVZ4 xenopus lae
31	688	10.1	1671	5	09N1J5	09N1J5 biomphalari
32	658.5	9.7	1368	13	08UW85	08UW85 paraliichthy
33	650	9.5	1418	13	093457	093457 scophthalmu
34	646.5	9.5	1369	13	08UW86	08UW86 paraliichthy
35	626	9.2	1358	8	073798	073798 xenopus lae
36	624.5	9.2	1472	5	09U5A8	09U5A8 bombyx mori
37	617	9.1	1412	13	08UW84	08UW84 paraliichthy
38	607	8.9	1245	13	09YGH8	09YGH8 scophthalmu
39	605.5	8.9	1418	13	08UW83	08UW83 paraliichthy
40	595.5	8.7	2144	5	09VD94	09VD94 drosophila
41	583	8.6	1371	11	09QVW4	09QVW4 ractus sp.
42	565	8.3	987	11	091YMO	091YMO mus musculu
43	563.5	8.3	1091	4	09UMQ4	09UMQ4 homo sapien
44	562	8.2	935	4	096135	096135 homo sapien
45	562	8.2	987	11	099WR2	099WR2 mus musculu

ALIGNMENTS

RESULT 1
018735
ID 018735 PRELIMINARY; PRT; 1259 AA.
AC 018735.
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JUN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ErB-2.
OS *Canis familiaris* (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "cDNA cloning of erB-2 from canine mammary gland";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008451; BA023127.1; -;
DR HSSP; P11362; IFGK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-1ike.
DR InterPro; IPR001245; Tyr_Pkinase.
DR InterPro; IPR004019; Furin-1ike.
DR Pfam; PF00757; Furin-1ike; 1.
DR Pfam; PF01030; pkinase; 1.
DR Pfam; PF02757; YLP; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TYKC; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 13798 MW; E37364D9CA4CD46 CRC64;

Query Match	89.9%;	Score 6129;	DB 6;	Length 1259;
Best Local Similarity	89.8%;	Pred. No. 0;		
Matches 1132;	Conservative 46;	Mismatches 75;	Indels 8;	Gaps 4;
QY	1	MELALCRWGLLLALLPGAASTOYCTGTDMKLRLPASPEHLDMRLHYGCCVQGNL	60	
DB	1	MELAAWCRWGLLLALLPGAASTOYCTGTDMKLRLPASPEHLDMRLHYGCCVQGNL	60	
QY	61	ELTYLPTNALSFLDIOEVGVYLIANOVQVLOGLRIVRGTOJFEDNYALAVDNG	120	
DB	61	ELTYLPANASLSFLDIOEVGVYLIANOVQVLOGLRIVRGTOJFEDNYALAVDNG	120	
QY	121	DLNNTTPTVGTASPGGLRELRJLSLTELKGGVLIQRNPOLCYODTILMKDIFKKNOLA	180	
DB	121	DLLEGIPAPGAAOGGLRELRJLSLTELKGGVLIQRNPOLCHOOTILMKDIFKKNOLA	180	
QY	181	LTLIDNRSRACHPSGPMCKSGRCWGESSEDCQSLTRVACGACRCGRLPTCCHEOC	240	
DB	181	LTLIDNRSRACHPSGPMCKSGRCWGESSEDCQSLTRVACGACRCGRLPTCCHEOC	240	
QY	241	AAAGCTGPKHSDCLACHFNHSGICEHCPALVTYNTDFESMPNDEGRYTFGASCVTACP	300	
DB	241	AAAGCTGPKHSDCLACHFNHSGICEHCPALVTYNTDFESMPNDEGRYTFGASCVTACP	300	
QY	301	YNYLSTDVSGCTLVPRLNQETADGQRCCKSKPARYCYGLMOYIKANSKFTIGIT	360	
DB	301	YNYLSTDVSGCTLVPRLNQETADGQRCCKSKPARYCYGLMOYIKANSKFTIGIT	360	
QY	361	ELFAGCKKIFGSLAFLEPESFDGDPASNTAPLOPEPOLVFETLEIIGYLYISAMPDLP	420	
DB	361	IOEFAGCKKIFGSLAFLEPESFDGDPASNTAPLOPEPOLVFETLEIIGYLYISAMPDLP	420	
QY	421	DLVSFONLOVITGRILHNGAYSLTQGLGISMGLRSIRELGGSLALIHNTLCPVHTV	480	
DB	421	NLSVFONLRVIRGRVLHNGAYSLTQGLGISMGLRSIRELGGSLALIHNTLCPVHTV	480	
QY	481	PMDOLFBNPHOALLHTANPREDECEVGEGLACHQOCARHCWGPPTQCVNOSOTLRQEC	540	
DB	481	PMDOLFBNPHOALLHTANPREDECEVGEGLACHQOCARHCWGPPTQCVNOSOTLRQEC	540	
QY	541	VEECRYLQGLPREYVNAHCLPCHPECOPONGSVTCFEPADQCAVAHYKDPFCVAC	600	
DB	541	VEECRYLQGLPREYVNAHCLPCHPECOPONGSVTCFEPADQCAVAHYKDPFCVAC	600	
QY	601	PSGVKPDLSYMPIMKFPDEEGACOPCPINCHSCVDLDDKCCAPRQASPLTISVSAVVG	660	
DB	601	PSGVKPDLSYMPIMKFPDEEGACOPCPINCHSCVDLDDKCCAPRQASPLTISVSAVVG	660	
QY	661	ILLVAVLVGVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNOAOMRLKXQYI	720	
DB	661	ILLVAVLVGVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNOAOMRLKXQYI	720	
QY	721	KANSKFTIGTEL-TYKGIWIPDGSNNKIPVAILKVLRENTSPKANKEILDEAVYAVGVS	779	
DB	721	KANSKFTIGTEL-TYKGIWIPDGSNNKIPVAILKVLRENTSPKANKEILDEAVYAVGVS	779	
QY	779	PVSRRLGICLTSTYQVLTQMLPYCCLDHRREHGRGSGDDLNNKCVQAKGMSYEDV	838	
DB	779	PVSRRLGICLTSTYQVLTQMLPYCCLDHRREHGRGSGDDLNNKCVQAKGMSYEDV	838	
QY	840	RLVRDLAARNVLVSPVHVKITDFGLARLLDIDETEXHADGCVPIKMALESILRRRP	899	
DB	840	RLVRDLAARNVLVSPVHVKITDFGLARLLDIDETEXHADGCVPIKMALESILRRRP	899	
QY	899	THOSDWSYGVTVMLMTFGAKPVDGIPAREIPLLLEKGEPLPPECITDVMYIMKCM	959	
DB	899	THOSDWSYGVTVMLMTFGAKPVDGIPAREIPLLLEKGEPLPPECITDVMYIMKCM	959	
QY	959	MIDSECRPFRELVSEFSRMAPDPORFVVIQNEDLGASPLDSTFYSLLEDDMDGLVD	1019	
DB	959	MIDSECRPFRELVSEFSRMAPDPORFVVIQNEDLGASPLDSTFYSLLEDDMDGLVD	1019	
QY	1020	ABEYLVPQGGFFCQDPAPAGAGMWHHRSSSTSSGGGDLTGLFSEBEAPRSPAPSE	1079	

DB	1019	ABEYLVPQGGFFCQDPAPAGAGMWHHRSSSTSSGGGDLTGLFSEBEAPRSPAPSE	1078	
QY	1080	GAGSDVFDGDLGMAKALQSLPTHDSPLQRYSEDDPTVPLPSTDTGVAPLITCSPOPEY	1139	
DB	1079	GAGSDVFDGDLGMAKALQSLPTHDSPLQRYSEDDPTVPLPSTDTGVAPLITCSPOPEY	1138	
QY	1140	VNQPPVRPQPSPREGRLPARPPGATLER-----AKTISGRKGVKVDVAFGAVENP	1194	
DB	1139	VNQPEVWPPPLALEGRLPSPRPAGATLERKTLSPKTLSPGKGVKVDVAFGAVENP	1198	
QY	1195	EYLTPQGAAPQHPHPAPFSPAFDNLVYWDODPEPERGAPSTFKGTPTAENPEYLGIDVP	1254	
DB	1199	EYLAPRGAAAPQHPHPAPFSPAFDNLVYWDODPEPERGAPSTFKGTPTAENPEYLGIDVP	1258	
QY	1255	V 1255		
DB	1259	V 1259		

RESULT 2

Q9QX70

PRELIMINARY; PRT; 1209 AA.

AC

Q9QX70;

DT

01-MAY-2000 (TRENBLrel. 13, Created)

DT

01-MAY-2000 (TRENBLrel. 13, Last sequence update)

DT

01-JUN-2002 (TRENBLrel. 21, Last annotation update)

DE

Epidermal growth factor receptor.

GN

EGFR.

OS

Rattus norvegicus (Rat).

OC

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX

NCBI_TaxId=10116;

RN

[1]

RP

SEQUENCE FROM N.A.

RC

STRAIN=FISHER; TISSUE=LIVER;

RX

MEDLINE=90258988; PubMed=2342466;

RA

Petch L.A.; Harris J.; Raymond V.W.; Blasband A.J.; Lee D.C.;

RA

Earp H.S.;

RT

"A truncated, secreted form of the epidermal growth factor receptor is

RT

encoded by an alternatively spliced transcript in normal rat tissue."

RMol. Cell. Biol. 10:2973-2982(1990).

RN

[2]

RP

SEQUENCE FROM N.A.

RC

STRAIN=FISHER; TISSUE=LIVER;

RA

Petch L.A.;

RN

submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.

RN

[3]

RP

SEQUENCE FROM N.A.

RC

STRAIN=FISHER; TISSUE=LIVER;

RA

Guttridge K.; Dawson T.L.; Earp H.S.;

RN

submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

DR

EMBL; M37394; A0114008.1; -

DR

HSSP; P1362; IFGK.

DR

Interpro; IPR000494; EGFR_L domain.

DR

Interpro; IPR000719; Euk_Kinase.

DR

Interpro; IPR002174; Furin-like.

DR

Interpro; IPR001245; Tyr_kinase.

DR

Pfam; PF00757; Furin-like; 1.

DR

Pfam; PF00069; Kinase; 1.

DR

Pfam; PF01030; Recep_L domain; 2.

DR

PRINTS; PR00109; TYRKINASE.

DR

ProDom; PD000001; Euk_Kinase; 1.

DR

SMART; SMO0261; Fy; 3.

DR

SMART; SMO0219; Tyrc; 1.

DR

PROSITE; PS00107; PROTEIN KINASE ATP; 1.

DR

PROSITE; PS0011; PROTEIN KINASE DOM; 1.

DR

PROSITE; PS00109; PROTEIN KINASE TYR; 1.

DR

ATP-binding; Receptor; Tyrosine-protein kinase.

SQ

SEQUENCE 1209 AA; 134891 MM; 96EEF7F6CC18773 CMC64;

Query Match 45.0%; Score 3065; DB 11; Length 1209;
 Best Local Similarity 49.6%; Pred. No. 4; 5e-224;

Matches 634; Conservative 164; Mismatches 363; Indels 118; Gaps 30;

```

QY 3 LAALCRMGILLALPPGA-ASTQVCTGDMKRLRLPASPETHLMDLRLHYGQCCVQGNLE 61
Db 15 LAALCAAG-----GALEEKVCOGTSNRNLTLGTFEDHFLSLQMFNNCEVLGNLE 66
QY 62 LTYLPPTNASTLFDLDIQGVGVLIANQVQVQLQSLRIVRGTLQFEDNYALAVLDNG 121
Db 67 ITTVQRNIDLSFLKTIQVAGVYLIANTVERIPLEMLQIIRGALYENYALAVLSN-- 124
QY 122 PLNNTPTVGASPGGLRELQRLSTLEIKKGVLLQQRNPOLCYODTILMKDIFPKNQQL 181
Db 125 -----YGNKKGLRELPRNQLQELLIGAVRSNNPILCNMETIMKDIV-QDVFLSN 175
QY 182 TLIDTNRS-RACHPCSPCKGSRGCBESSDQSLRTVCAAGCA-RCKGPLPTDCHEQ 239
Db 176 MSMQVQRLTGCCPKCDPSCPNKSCWGRGEEQCQLTKICAQCCSRRCRGRSPSDCHNQ 235
QY 240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMNPGRRTFGASCYTAC 299
Db 236 CAAGCTGPRSDCLVCHRFREDATCTKOTCPMLLYNTTQMDVNPBGKYSFGATCVKGC 295
QY 300 PYNLTSTDVGSCTLVCPLANQVTAEDGTQRCCKSPKARVCYGLMQYIKANSKFIGI 359
Db 296 PRNYYVTDHGSQVACGPDYEV-EEDQVSKCKCDGPRCKVCMGIGIGBPK-DLISINA 353
QY 360 TELF-FACCKKIFGSLAPFESFDGPDASNTAPLQEPOLQVETLEETIGYLYTSAMPDS 418
Db 354 TINKHFKYCTAISDDLILPVAFFGDSFTPTPDPRELEILITVKEIGFLLIQAMPEN 413
QY 419 LPDLISVFNQLQVIRGRILHNGAYSLTQGLISWLGSRSLREKSGSLALHHTHLCFYN 478
Db 414 WTDLHAFENLEIIRGRKTHQGFSLAVGNLITSGLRSLSKESDDGVITISGRNLCYAN 473
QY 479 TVPMDQLEFRNHQALLHTANPEDECVGEGIAHQICARGHCWGPPTQCVNCSQFLRQ 538
Db 474 TINKKTLFGTQNKTKIMNNAEKDCATNHCVCPLCSSBGCGPPTDVCQNVSRGR 533
QY 539 ECVEECRYLQGLPREYVNAHRCULCPHECQPNQSVTCFEPEDQVCAAHYDPFCYA 598
Db 534 ECVDKCNILBEPREFENSECICQHEPCLPQTNKTLCTGRGPNCKICAHYVDGHCWK 593
QY 599 RCPGKVRKDISYPMWKFPEEGACQPCPINTCHSCVDLDDKCPAPORASP-LTSIVA 657
Db 594 TSPSGIMGENTTL-VMKFADANNVCHLCHANCTYGCAGPGLKNC-QQBPGRIPBIAAG 650
QY 658 VSGILLVYLVGVVGT-LIKRQCKIKRYMRLLQETLVEPLTSGAMPQACQRIILK 716
Db 651 IVGGLFTIV-VALGICLFMRKRLQVARKTLRLQLERELVEPLTSGEAPNQAHILIK 709
QY 717 EQYIKANSKEFIGTEL-TVYKGIWIPGGENVKIPVAIKVLRENTSPKANKETIDEYVNA 775
Db 710 ETEFK-KIKVLGSAFQTVYKGLMIPGEKVKIPVAIKELREKITSKANKELIDEYVNA 768
QY 776 GYSGPYVSRLLGTLSTVOLVQLMPYGCGLDHYENRGLCSQDLLNWCQIAKMSY 835
Db 769 SYDNFHVCRLLGLCTSTVQLITQLMPYGCGLDYVREHKDNISQYLNMWCQIAKMSY 828
QY 836 LEDVRLVHRDLAARNVYVKSPPNVKIFDFGLARLLIDENEYVADCGKVPIMWLESL 895
Db 829 LEDRVLVHRDLAARNVYVKTPOVKITTDRLALCLGAEKRYHAGGVKIPIMWLESL 888
QY 896 RRRPTQSDVWSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIM 955
Db 889 HRIYTHQSDVWSYGVTVWELMTGSKPYDGIIPASEFISILEKGERLPQPICTIDVYIM 948
QY 956 VKCMIMISECRPRELIVSEFSYARDPQRFVYIQ-NELGSPSLDSTFFYSLLDDDM 1014
Db 949 VKCMIMIDASRPKFRLEILIEFSKARDPOKIVLYIQDDEMHLPSPPTDSFYALMEEDM 1008
QY 1015 GDLYDAEYVLPQGGFFCPDPABAGAGMWHRRSSSTRSGGDLTLGLPESEEPAPSP 1074
Db 1009 EDVYDADEYVLPQGGFF-----NSPST-----SRTF 1034

```

```

QY 1075 LAPSEAGSDVFDGDLGKAAGLQSLPTHDPSPLQRYSEDPYPLPSET--DGYVAPLT 1132
Db 1035 LLSLSIASNSN-----STVACTINRNGSCRYVEDALQYSSDPTSVLTEDNIDTFL---- 1086
QY 1133 CSPQPEYVNPQDVAPQPPSPREGFLPARPAGATLEBAKTLSPGKNGVXDVAFAGAVE 1192
Db 1087 --PVPEYINQ-SYPRKPRAGSVQNVYHNQPLHP-----APQRDLHYQN--PHSAVS 1133
QY 1193 NPEYL-TPQGAAGQHPHPPAFSPFNLYVMQ-----DP-----PERGAPPS 1235
Db 1134 NPEYLNTAQ-----PCLSSGFSALMIQKSHQMSLDNPYQODFFPEAKPNG 1184
QY 1236 TFKGPTAENPEYGLDVP 1254
Db 1185 IFKG-PTAENPEYLRVAPF 1202

```

RESULT 3

```

Q9EP98
ID Q9EP98 PRELIMINARY; PRT; 1210 AA.
AC Q9EP98;
DT 01-MAR-2001 (TRENBERL 16 Created)
DT 01-MAR-2001 (TRENBERL 16 Last sequence update)
DT 01-JUN-2002 (TRENBERL 21 Last annotation update)
DE Epidermal growth factor receptor isoform 1.
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearse R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maile N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Schehl C., Pearse R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maile N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF275366; AAG28045.1; -.
DR EMBL; AF275364; AAG28045.1; JOINED.
DR EMBL; AF275365; AAG28045.1; JOINED.
DR EMBL; AF275367; AAG24386.1; -.
DR HSP; P11362; IRGK.
DR MGD; MG1195294; Egfr.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002299; Ser_Thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FUF_5.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TYKc; 1.

```

DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN_1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP_1.
 DR PROSITE: PS0011; PROTEIN_KINASE_DOM_1.
 DR PROSITE: PS0109; PROTEIN_KINASE_TYR_1.
 KW ATP-binding; Receptor; Transferrase.
 SQ SEQUENCE 1210 AA; 134840 MW; 62CD021C9DE32E18 CRC64;

Query Match 44.5%; Score 1036; DB 11; Length 1210;
 Best Local Similarity 49.1%; Pred. No. 7.3e-222;
 Matches 625; Conservative 166; Mismatches 369; Indels 114; Gaps 27;

QY 11 LLLALLPGAA--STQVCTGIDMKLRPASBETHLMDLRHLYOGCOVVOGNTLETTYPTN 68
 DB 14 LTTALCAAGALEEKVKVCGTSNRLTQGTEDHPLSLQRMWYNNCEVVLGNLETTYQRN 73
 QY 69 ASLSFLQDIQEVQVYLLAHNQYRQVPLQRLIRVGTQLFEDNALAVLDNGDFLNTTP 128
 DB 74 YDLSFLKTIQEVAGVYLLALNTERIPLENLQILRGNALYENTYALALSN----- 124
 QY 129 VTGASPGGLRELQLRSLTEILKGVLLIQNPQLCYQDTILMKDI----FKNNQLALTLI 184
 DB 125 -YGTNRTGLRELPMENLDELILGAVFSNNPILCNMDTIQMRDLYQNVFMSNMSMDL--- 180
 QY 185 DTRNSBACHPCSPMKCGSRGWSESDQSLTRIVYAGGCA-RCKGPLPTDCCHGQCAAG 243
 DB 181 -QSHPSCKPCDPCPNCSGWGGGECNQKLTIXICAQCCSHRCGRSPSDCHNQCAAG 239
 QY 244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTPESMPREGRTFGASCYACPYNY 303
 DB 240 CTGPRESDCLVQCKQDEATCKDTCPRLMLYNPTTYQMDVNEBSKISFGATCYAKCPRNY 299
 QY 304 LSTDVGSCTLVCPRLHNOETVADGTQRCCKSCKPCARCYVGLQMVYIKANSKFIGITELE 363
 DB 300 VVTDBGSCVACGPPYEV-EEDGIRKCKCKDGCRCXKCNIGIGERK-DTISINATNIK 357
 QY 364 -FAGCKIFGSLAPFBESFDGDPASNTAPLQPELOVETLEETGLYLYISAMPDSDL 422
 DB 358 HFKYCTAISGDLHILVAFKQGSFTRTPDLRELEIKYKELTGFLLLQAMPDMNTDL 417
 QY 423 SVFONLQVIRGRILHNGAYSLTLOGISLWGLRSLRELGSGLALIHNNHLFCVHTVP 482
 DB 418 HAFENLEIIRGTRKHQGFSLAVVGLNLTSLGSLKSIKISGDVITISGNMNLCAANTIN 477
 QY 483 DOLFNRHQALLHTANREDECEVEGELGACHQLCARHCWGPGPQCNCQOFLRQCEVE 542
 DB 478 KKLFGTPNQKTKIMNNRAEKCKCAVNHVCNPLCSGECGWGPEDPCVSCCNVSRGECVE 537
 QY 543 ECRVLQGLPREYVNAHCLPCHPEGOPONGSVTCFGEADQCVCAHYKDPFCVACRPS 602
 DB 538 KGNILEGEPREFEVSECIQCHPECLPQAMNITCTGAGPNCICQAHYIDGPHCVKTCRA 597
 QY 603 GYKPLSLTMIPIWKPFDEGACQPCPINCTHSQVLDLDDKCPAEQRPASPLTISAVVGLI 662
 DB 598 GIMGENNTL-VWKYADANNVCHLCHANCTYGCAPRGQCGEVPWSPGKIPSIATGIVGL 656
 QY 663 LVVVLGVVFGI-LIKRQOKIRKTYMRLLOETELVEPLTPSGAMPQAOAMRLIXQYIK 721
 DB 657 LEIVY-VALGIGLPMRHRHIVRKATIRLLDOERLVEPLIPSGAPQAOALRLIKETERK 715
 QY 722 ANSKFIGITEL-TYVKGIMIPDGNVNI PVAIKVLRNTSPKANKETLDEAVYVAGVSP 780
 DB 716 -KIKVLGSGAGFATYVKGIMIPDEGKVKIPVAKIKELRATSPKANKETLDEAVYVAVDNP 774
 QY 781 YVSRLLGI CLTSTVQLVQMPYCGCLDHYRENGRGLGSDLLWMCQILKMGSYLEDVR 840
 DB 775 HVCRLLGICLSTVQLVQMPYCGCLDHYREHNDNGSGLLWMCQILKMGSYLEDVR 834
 QY 841 LVHRDLAARNLVKSPNVKXITDFGLARLLDIDETEHAGGKVPPIKMALESILRRFT 900
 DB 835 LVHRDLAARNLVKTPGVKITTDFGLAKLGAEBKEVHAAGKVPPIKMALESILRRIT 894
 QY 901 HOSDVMYSYGVTVWELMTFGAKFYGTIGARELPDLLEGELPOPPCTTIDVYIMWKCM 960

DB 895 HOSDVMYSYGVTVWELMTFGSKPYDGIASDISLLEKGEPLPQPICTIDVYIMWKCM 954
 QY 961 IDSECRPFRELIVSEFSMARDPQRFVVIQ-NEDLGASPLDSTFYRSLLEDMDGLVD 1019
 DB 955 IDADSRPKFRELILEFSMARDPQRFVVIQODERMHLPSTDSFYALNDEDEMDVD 1014
 QY 1020 AEEYVLPQGGFFCPDPAAGACGMVHHHRSSSTSGGGDITLGLFSEBAPSPAPSE 1079
 DB 1015 ADEYLTLPQGGFF-----NSPST-----SRTPLLSL 1040
 QY 1080 GAGSDVFGDGLGMAKALQSLPHNDESPLORYSEDPVLPSET--DGVAFLPSCPOP 1137
 DB 1041 SATSN-----NSTVACINRNGSCRVKEDAFLORYSSDPTGAVTEDNIDATL-----PVP 1090
 QY 1138 EYVNPDPVFPQPSFESGPLPAAPAGATTLERAKTSLPGKNGVYKVFAGAVENPEYL 1197
 DB 1091 EYVND-SYKRRPAGSVQNPVYHNPPLHP-----APGRDLHYQN--PSNAGNPEYL 1139
 QY 1198 -TPQGAAPQHPHPAPFAPFADNLVYWDQ-----DP-----PERGAPSTFGKT 1240
 DB 1140 NTAQ-----PTCISGFSNPAIIOKGSHOWSLNDPQODFFPKETKNGIFKG- 1189
 QY 1241 PTAENPEYLGIDVP 1254
 DB 1190 PTAENAEYLRVAPP 1203
 RESULT 4
 Q9YH40 PRELIMINARY; PRT; 1.65 AA.
 AC Q9YH40;
 DT 01-MAY-1999 (TRENBLrel. 10, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, last annotation update)
 DE Receptor tyrosine kinase proto-oncogene.
 GN XMRK.
 OS Xiphophorus xiphidium.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC Acanthomorphi; Acanthopterygii; Percormorpha; Atherinomorphi;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OX NCBI_TaxId=8086;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIO PURIFICATION;
 RX MEDLINE=98241172; PubMed=9582016;
 RA Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,
 RA Altschuld J., Scharlt M.;
 RT Activation of the xmrk proto-oncogene of Xiphophorus by
 RT overexpression and mutational alterations.";
 RL Oncogene 16:1681-1690(1998).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIO PURIFICATION;
 RA Scharlt M.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U53471; AAD10500.2; -.
 DR HSB; P11362; IFGK.
 DR InterPro; IPR000345; CytC_heme_bind.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_Dkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR001245; Tyr_Dkinase.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR PRINTS; PS00109; TYRKINASE.
 DR ProDom; PD000001; Euk_Dkinase; 1.
 DR SMART; SM00261; FU; 3.
 DR SMART; SM00219; TYRK; 1.
 DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN_1.
 DR PROSITE; PS00343; GRAM_pos_ANCHORING; UNKNOWN_1.

DR PROSITE, PS00107; PROTEIN KINASE, ATP, 1.
 DR PROSITE, PS50011; PROTEIN KINASE, DOM, 2.
 DR PROSITE, PS00109; PROTEIN KINASE, TYR, 1.
 KW ATP-binding; Kinase; Tyrosine-protein kinase.
 SQ SEQUENCE 1165 AA; 129614 MW; 7f7EE38D8771A4E CRC64;

Query Match 39.0%; Score 2656; DB 13; Length 1165;
 Best Local Similarity 44.9%; Pred. No. 5,6e-193;
 Matches 574; Conservative 163; Mismatches 390; Indels 150; Gaps 33;

1 METALCRWGLLALPPG-AAST---QVCTGDMKRLPASPETHLDMRLHVGCGV 55
 4 LELLE-----LILLILSLGRCSTDPDRKVCQGSNQM---LDNHYLKMKKQSGCNV 56
 56 VQGNLELYLPTNASLFLQDIQEVGVYLLAHNQVQVPLORLRIYRGTLFEDVYALA 115
 57 VLENLEITYQENQSLSTQSIOEVGVYLLAHNEVSTIPLVNLRLRGNLVEGNFTLL 116
 116 VLDNGDPLNTPVYGASPGGLREQLRSJLEILKGSVLLQRPOLCYODTIIMKDI FHK 175
 117 VMSNYQK-NPESP--DYVQVGLKQQLSLNLEILSGGVKSHNPLLCNVETIMWMDIVDK 173
 176 NNOALTLIDTRSRACHPCS PMCKGSRGWESSEDSQSLTRTVCAQAGC-AQCKGPLPD 234
 174 TSNPTNWLIPAFERQCKDPCGVNCSMAFGCHQKTKLCAEQCNRRGRGRPID 233
 235 CCHQCCAAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDIFESMPNDEGRYTFGAS 294
 234 CCHQCCAAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDIFESMPNDEGRYTFGAS 293
 295 CVTACPVYVLTQVSGCTVPLHNOEVLADGOREKSKCAVVCGLM---QYI 350
 294 CVKCPSNVYVTE-GACVRSASAMLEVD-ENGRKSKCPDGVCPYVCGIIGISLNTI 351
 351 KANSKFTIGITELEFAGCKIIFGSLAPLPESPDDPASNTAPLOEOLQVETLEBITGYL 410
 352 AVNSTNIG---SFSNCKTINGDIIILNRSFEGDPHYKIGPMDPEHMLLTVKEITGYL 407
 411 YISAPDLSLPDLSFQNLQVIRGILHNGAYS-LTIQGLSILGIRSLREIAGSLALIH 469
 408 VIMMPEMNTSLSFQNLQVIRGILHNGAYS-LTIQGLSILGIRSLREIAGSLALIH 467
 470 HNTLCEVHTVPMQDLFRNPHQALLHTANRPEDECVGEGLAQHLCARHCGPGPTQCV 529
 468 NTPQLRYASTINMRRLPRSEDQSIENDART-----ENQTCNNESEGGCGPGTQCV 520
 530 NCSQFLGQCEVEECVYLGCLPREYVNAHCLPCHPECCOPONGSVTCFEPREADQVACAH 589
 521 SCHVDHGRGVASCNLLQGEPRQAVDGRCVQCHQECVLQDTSLTCYGPGRANCSKCH 580
 590 YKDPFCVACRPSGVKXDLISYMIKFPDEEGACOPCPINCTHSCVLDLDDKCPAEQAS 649
 581 FQDQPCIPRCPHGMIGDGTLL-IMKXADMGGCQCHQCTQGGSGPGLSGCRD-IVS 638
 650 PLTSIVSAVVGILLVVLGVVFIILIKRQOKIRKXTMRRLQETELVEPLTPSGAMENQ 709
 639 HSLAVGLVSGLLITVALLIYVLLRRRRIR-RKXITRILLQEKELVEPLTPSGAENQ 697
 710 AQMRILKEQYIKANSKFIETEL-TYRKGIWIPDGNVNIPIAIVKLEBNTSPKANKEL 768
 698 AFIRILKETEFK-KDRLVLSGAGFVYKGLMPDGENIRIPAVIVLEBASPKVQCEVL 756
 769 DEAYVAVGSSPYVSHLLIGLITVQLTQCLMPYGCGLDHYRENRGLSGDILLNMQ 828
 757 DEAYVAVSDHPRVGLLIGLITSAVQLTQCLMPYGCGLDHYRQOEIRICGMLNMQCVQ 816
 829 IAKGMYLEDVLYNHDLAARVLYKSPNHVKITDPGLARLLIDIDETVYHADGGKVPILK 888
 817 IAKGMYLEDRHLYNHDLAARVLYKSPNHVKITDPGLARLLIDIDETVYHADGGKVPILK 876
 889 MALESILRRRFTHQSIVMSYGVTVMLMFGAKPPDGIIPAREIPDLLEKGRLLPQPICT 948
 877 MALESILQWITYHQSIVMSYGVTVMLMFGAKPPDGIIPAREIPDLLEKGRLLPQPICT 936

QY 949 IDVYIMYKCMWIDSECRPRRELVSERKAPDQRPVYIQNEDLQASPLDSTFYRSL 1008
 DB 937 IEVYIILKCMWIDPSSRPRELVLGERSQWAPDSRYLVQ---NLPSFSDRLRSL 993
 QY 1009 LEDDMGDLVAEEYLVQCGFFCPDPAPAGGVNHRHSSSRSGGDLTLGLESEE 1068
 DB 994 LSSDD-DVDADELTL-----RIRKIN-RQGS----- 1018
 QY 1069 EAPRSLPAPSGAGSDVFDGLGMAKGLQSLPTHPDPSLPORYSEDPTV-PLPSETDGY 1127
 DB 1019 ---EPCIPENGH-----FVRENSIALRYISDPTQNALEKLDGH 1054
 QY 1128 VAPLTCQPEYVNOQDVRQP-----PSPR-----GPLP-AAPRAGATLERAKTIS 1174
 DB 1055 -----EYVQPGSEYSSRLSDIYNRYEDLTGWSGVLSQGEARTNRPPEYLN 1104
 QY 1175 PGRGVNDVFAFGAVENPEYLPQGAAPQHPHPAPAFDNLVYWDQDPPRGAP 1234
 DB 1105 TNQNSL---PLVSGSSMDPDY---QAG-----YQAF-----LPQTGALT 1139
 QY 1235 STPKGTAPENPEYLG 1251
 DB 1140 GNGMFLPAENLELYGL 1156

RESULT 5

ID Q9M6F6 PRELIMINARY; PRT; 1137 AA.
 AC Q9M6F6;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Receptor tyrosine kinase (Fragment).
 GN BR34.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OK NCBI_TaxID=9031;
 RN [1]
 PP SEQUENCE FROM N.A.
 RC TISSUE=HINDRAIN;
 RX MEDLINE=99263203; PubMed=10328884;
 RA Dixon M., Lumsden A.;
 RT "Distribution of neuropilin-1 (nrg1) and eirb4 transcripts in
 RT embryonic chick hindbrain".
 RL Mol. Cell. Neurosci. 13:237-258(1999).
 DR HMBL, AF121963; AAD31764.1; -.
 DR HSSP; P11362; 1FGK.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001368; TNFR_c6.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR InterPro; IPR004019; YLP_mocif.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF01030; Recep_L_domain; 1.
 DR Pfam; PF02575; YLP_2.
 DR PRINTS; PRO0109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 3.
 DR SMART; SM00219; TYKIC; 1.
 DR PROSITE; PS00107; PROTEIN KINASE, ATP, 1.
 DR PROSITE; PS50011; PROTEIN KINASE, DOM, 1.
 DR PROSITE; PS00109; PROTEIN KINASE, TYR, 1.
 KW Kinase; Tyrosine-protein kinase.
 FT NON_TER 1
 SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F8D0C4F CRC64;

Query Match

38.4%; Score 2615.5; DB 13; Length 1137;

Best Local Similarity 45.9%; Pred. No. 6,66-190;
Matches 524; Conservative 174; Mismatches 362; Indels 81; Gaps 28;

```

QY 161 LCYQDTILMKDIFHNQNLALTLITNSRACHPCSPMKCKSRGWSSSEDQSLRTVC 220
Db 3 LCFADTIIHMODIVRNPMASNFTLVPTNSSGGRHNSCTG-RWGPTEHNCQTLLTITVC 61
QY 221 AGGC-ARCKGPLPTDCCHEQCAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTF 279
Db 62 AECCDGRCYGYPVSDCCHEQAGGCGSKDTCACAMNFNSGACVYQCTQPTFYVNTPTTF 121
QY 280 ESMNPBGRITFGACCVACAPYNTSTVSGCTLVCPHLNCEVTAEDGTORCEKSRPCA 339
Db 122 QLEHNNAKYTYGACVCKCPNHFV-VSSSCVACSSKMEV-EENGIKCKRCPCTDIP 179
QY 340 RVCYGLQMAYIKANSKFGITELF-PAGCKKIFGSLAFPESPGDASNTAPLOPQ 358
Db 180 KACDGIQTGSL-VSAQYDSSNIDKFNCTKINGNLIFVTGIGHDYHTTAAINPEKLN 238
QY 399 VETLEITGVLYISAMPDSLPLSVFQNLGVIRGLIHNGAVSLTQGLGISMGLRL 458
Db 239 IFQVREITGYLNTIOSMBENMTDFRVSNTLVITIGRALYSGLSLLIKQGITSLQFSL 298
QY 459 RELGSGLALIHNNTHLCFVHTVPMDOLEFRNHQALHTANRPEDECGESLACHQCAR 518
Db 299 KOISAGNIYITDNSMLCYHTVNMSTLESTPSQKTVIHRKKKANTCADMVCNELCSSD 358
QY 519 HGWGPPTQCVNCSQFLRGOECVECEVYLOGLPREVYNAHCLPCHRECP-ONGSVTCE 577
Db 359 GCMGPDQCLCKFRIGRTICISNLYDEFRFANGVCMEDCPQCKMEDNMITCY 418
QY 578 GPEADQVACAHYDPRFCVACSGVYKPLSTYPIAKFPDEBACOPCINCTHSCVDL 637
Db 419 GPBDHCTCKCFHFDGPRVCCKCDGLGANSF--IKYVADRECHPCPHNCTOCGRCP 476
QY 638 DDKG-----PAQCRASPLTSIVSAV--GILLVAVLGVFGILIKRQCKIRXY 685
Db 477 ASHDCIYVWTRQSTLPPHAR-TL--IAGVIGGLIYIMGLTFAVYRRKSRK-KKR 532
QY 686 TMRILQTEIVEPLTSGAMPNQAQRILKEQYIKANSKFGITEL-TYKGIWIPDE 744
Db 533 ALRFL-ETELVEPLTSGTAPNQAQRILKETELK-RVVLVLSGAFGIYKGIWIPDE 590
QY 745 NVKI PVAIKVLRNTSPKANKELIDEAYVAGVSPVSLTICLSTVOLVTOUMPG 804
Db 591 TVKIPVAKIKINETGKRANVEFDELIMASDHPLVLLGVCLSPITQLVQLMPHG 650
QY 805 CLDHYVNRGRGLSGOULLWKCQIAKMSYLEDVRLVHBDLAARVLYKSPHYKITDF 864
Db 651 CLDHYVNRGRGLSGOULLWKCQIAKMSYLEDVRLVHBDLAARVLYKSPHYKITDF 710
QY 865 GLARLLIDETEVYADGVPIKMMALLESILRRRFTHOSVWSYGTWELMTFGAKPD 924
Db 711 GLARLLIDETEVYADGVPIKMMALLESILRRRFTHOSVWSYGTWELMTFGAKPD 770
QY 925 GIPAREIPDLLEKGERLPPICTIDVYMIYKCMWIDSECRPREFRELSEFSRMAADPQ 984
Db 771 GIPAREIPDLLEKGERLPPICTIDVYMIYKCMWIDSECRPREFRELSEFSRMAADPQ 830
QY 985 RFVYIQNEE-LGPASPLDSTFYRSLIEDDMGDLVAEEYLVYQGFPPDPAPGAGV 1043
Db 831 RYLVIQDQDKMLPSPNDSKFFQVLDEEDLECMMAEEYLVY-QAFNIPPIYTSRTIR 889
QY 1044 HHRHRSSTSGGGDLTLGLEPSEEAPRS--PLAP-SEGAGSDVVDGDLGGAAGKQS 1100
Db 890 DSNRNGFYVRDGGVAAEQGV-FMPYRAPGCIIPBAVAGATAEIEDCCNGTLRKQYA 948
QY 1101 LPHDSEPLQRYSEDPPVPLS-----ETDGYVAPLTCQPEYVNOVDYRPPSPSPR 1153
Db 949 TLAKEDSSTQRYSDPVLPIPERVIGELDEDDGYMTPMDRCKRTYLVNVEENPFVSRK 1008
QY 1154 EGRLPAA-RPAGATLERAKTLPGRKGVVYVF-----AFGAVENPEVLTQGGAA 1204

```

```

Db 1009 NCDDLOAVNPEYHN-----APNGCPKAEDEYVNEPLVNTFANTLENABYL----- 1054
QY 1205 PQHPPEAFSAFDNLYYWDGDPERGA--PPSTFKGPT-----AENPEY 1248
Db 1055 -KNLPERAKKAFNDPVMHNSLPFRSLGHQPDVLQGYSTKYFYKQNGRIPIVAENPEY 1113
QY 1249 L 1249
Db 1114 L 1114

```

```

RESULT 6
P79754
ID P79754 PRELIMINARY; PRT: 1328 AA.
AC P79754;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Brb3.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99177347; PubMed=10077531;
RA Gellera X., Brenner S.;
RT Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
RT rubripes.
RT Genome Res. 9:251-258(1999).
DR EMBL: AF056116; AAC34391.1; -.
DR HSSP: P11362; IFGK.
DR InterPro: IPR000494; EGFR_L domain.
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_Pkinase.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; Pkinase; 1.
DR Pfam: PF01030; Recep_L domain; 2.
DR ProDom: PD000001; Euk_Pkinase; 1.
DR SMART: SM00261; FU; 3.
DR SMART: SM00219; TYRK; 1.
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KV ATP-binding; Transferase.
SQ SEQUENCE 1328 AA; 148613 MW; A133039258B647E9 CRC64;

```

Query Match 32.6%; Score 2225; DB 13; Length 1328;
Best Local Similarity 39.6%; Pred. No. 4,3e-160;
Matches 511; Conservative 156; Mismatches 424; Indels 200; Gaps 34;

```

QY 9 WGLIALLP--GAASQ-----VCTGDMKLLPASPETHLMDLHLVGGCQVAGNLEI 62
Db 4 WRLIMVARSRLAASASOTQBAVCPETQGLSSTGQENQVYLNDRKXGCIINGNLEI 63
QY 63 TYLPNASSFLDIOEGVGYLIANNOVROYPLGRIVRGTOLEFEDNYALAVDNDP 122
Db 64 TQIESWDFEFLLKTIRETVGYLIANNHFOEIPLOGLRIVRNSLYERFALSVLN--- 120
QY 123 LNNTTYVYGASPGCLRELORLSTELIKGVYLIQRRPOLCYQDTILMKDIFHNQNLALT 182
Db 121 ----YKDG--PSGLQLQELKMLTETLDDGVQI INKKYLRYPWYWRDII-RNNDABIE 173
QY 183 LIDTNSRACHPCSPMKSGSRGWSSSEDQSLTRTVAGG-ARCKGPLPTDCCHEQCA 241
Db 174 IQFNGRGYCH--KSC-GNYCMGKGKQCQLLTITVCAQPCNDSCFSTSRDCHIECA 229
QY 242 AGGTGKSHDCLACHFNHSGICELHCPALVYNTDTESMNPBGRITFGACSVTACPY 301
Db 230 AGCKGPLDTCACARLFDNSGACVPOCQTLIYNKQTFQMETNPNKQYQYSGISVSCPT 289

```

```

QY 302 NYLSTDVSGCTLVCPBLNQEVTADDTOR-CEKSKRCAQVVCGLMQUYIKANSKFIGIT 360
DB 290 HFV-VDGSSCVSVCEPDMEV--ERGSQROCELSGCLPVCYCEGTAE---QROVSS 342
QY 361 ELE-FAGCKKIFGSLAFLPESFDDPASPANTAPLOPELOVFEETLEITGYLYISAMPDL 419
DB 343 NIDSFINCTKIOGSHFLVGLIGDDPRKNNPDLAKLEFVRIVREITOLIMIGMPKEL 402
QY 420 PDLVFNOLQVIRGILHNGVSLITLOGLISWGLRLSELGSLALIHNTLCEVHT 479
DB 403 NDLVSFSLITIOGSLERKPSLWMRIPLTSLGLRLSELGSLALIHNTLCEVHT 462
QY 480 VPMQDLFRNPH-QALLHTANPEDECEGELACGTCARHCGPPTOCVANCSCGFLRQ 538
DB 463 VMMTOLFRGSRVRANSLSNRPNAECVADGRVCDPLCSGCGCPDPDCLSCRYNSRIG 522
QY 539 ECVEECRVLOGIPREYVNAH-CLPCHPECPONGSYTCGPAPADOCVACAHKDPFCV 597
DB 523 TCVAGCHFNSGIPREFAGLVNCVACHPECKPOTGKASCCTGPADOCMACTKFRDGPYCM 582
QY 598 ARCPGVKPDLSYMPIMKFPDEGACOPCPINCTHSCVDLDDKGCAPAGRASPLTSISA 657
DB 583 SSCPAGVA-DGEKGIIFKFPVRECHCECHNCTOGSGSGGLND---LEAARLITSSQ 638
QY 658 VVGLLVVLCVVF-----GLIKRQOKIRKYMRLLOETELVEPLUTSGAMPNCA 710
DB 639 ITGIALGVPAGLIFCLVLPFLGMLYHRGLAIRKKAMRRRLYESGESFEPLGP-GEKGTGV 697
QY 711 QMRILKEQYIK---ANSKFIGITELTVYKGINVPDENVKIIVALKVIRENTSPKANK 766
DB 698 HARILKPSDLKIKRPLSGVFC---TVSKGFNIPBEETKIVALKTIQDSGKOTFE 753
QY 767 ILDEAYVAVGVSGVPSRLIGTCLTSTVQVLTQVLMYPCGLDHVRENRLSGQDLNWC 826
DB 754 ITDHLISWGSLDHPIYVALLIGICGTCLQVLTOLSSGSLHEIRHOKTSLDPORLNWC 813
QY 827 MOIAKGVSYLDEVLVHRDLAARNVLKXSPNHYKITPFGARLLIDETEHYADGKVP 886
DB 814 VOIAKGVSYLDEHVRHKNLAARNILKNDYOYOISYGVADLLYDDKITYYSEKTP 873
QY 887 KMALESTLRRRFTHQSDVSVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPORPI 946
DB 874 KMALESTLRRRFTHQSDVSVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPORPI 933
QY 947 CTIDVYIMVYKCMWIDSECRPRELSEFSRMAPRDPQRFVIVQNDLGPASPLDSTFVR 1006
DB 934 CTIDVYIMVYKCMWIDSECRPRELSEFSRMAPRDPQRFVIVQNDLGPASPLDSTFVR 980
QY 1007 SLEBDDMGDLVDAEYVLPQGFCCPDPAPGAGGVNHHRRSSSTRSGGDLTGLEPS 1066
DB 981 --EDSGMGBFL-----RGSER---GLLEADLEED 1005
QY 1067 EEBARSPPLARSEGSDVFDGLMG--AAKGLSLPHDPSPLQ-----RYS 1113
DB 1006 EBE-----GLDPRATSLDPSRSMWSTSPQINSYVMMTQLRND 1044
QY 1114 EDPVPLPSETDGVAPLTCGPP-EXVNO-----PDVRQPPSPREGPL 1157
DB 1045 ---FAVSGQGHIGLPMSPFVDITRQMYQSRSLSVTLTPRASFRRSSREALEC 1098
QY 1158 -PAARPAAGATLERAKTISPCKNGVNVVFAFGAVNPELTPOGGAAQPHPPAFSPA 1216
DB 1099 EDGAGCAGIFRVR-----FSGERGN-----PQGG----- 1122
QY 1217 PDNLVYWDQPPPERGADPSTFKGTPTAENPE 1247
DB 1123 -----QQRKLSTASSPSSFKTWADEDE 1146

```

RESULT 7

098IH9
ID 098IH9 PRELIMINARY; PRT; 1433 AA.
AC 098IH9;

```

DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
OS
GN Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_TaxID=7165;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=SU4;
RA Lycett G.J.;
RT Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor.
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301655; CAC35008.1; -
DR HSP; P11362; 1FGK.
DR Interpro: IPR000345; CysC heme bind.
DR Interpro: IPR000484; EGFR_L_domain.
DR Interpro: IPR000719; Euk_kinase.
DR Interpro: IPR002174; Furin-like.
DR Interpro: IPR002290; Ser_thr_kinase.
DR Interpro: IPR01245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; kinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PRO0109; Recep_L_domain.
DR ProDom; PD000001; Euk_kinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS00219; TYKC; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON_TER
SQ
SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;
Query Match 28.9%; Score 1971.5; DB 5; Length 1433;
Best Local Similarity 32.1%; Pred. No. 9,4e-141;
Matches 465; Conservative 198; Mismatches 398; Indels 387; Gaps 39;
QY 26 CCGTDMKLRLPASPTLDMRLHLVQSGCVVQSLLELYFTNALSFLDDIQEVGYVL 85
DB 1 CIGTGRMSVPMANREYHYNLRDRYTNCTVVDGNLEITWIONITDLNFIHREVTGYVL 60
QY 86 IAHNVQVPLQRLRIVGTOLF-----EDNYALAVLDNGPLNNTTPTVGSPGGLREL 140
DB 61 ISLYDLPIYLRLRIIIGRTTFKLNKKEEYGLFV-----SFSMNTL 104
QY 141 QLRSLTEIKGVLIQRPOLCYODTIMKDI-FHKNQALALTLIDNRSRACHPSPMC 199
DB 105 ELPALRLDIGSGVGFNNYNNLCHMSIMWEILLAPQTSQYTFNFSSPRVCPGHPSC 164
QY 200 KGSRCWGESSEDCSLTITVACAGCA--RCGRLPTDCCHEGCAAGCTGKSHDCLACH 257
DB 165 EVG-CWGBAHNCQAFSLKNSPQSGRCGPRPRCHLFCAGGCTGTGSDCLACKN 223
QY 258 FHNSGICELHCPALVYNTDTFESNPNEGRYTFGASCVTACAPYNYLSDVSGCTLVCP 317
DB 224 FYDDGVCKQECPPMCIYAPTYVFEPPNDGCKAYAGACVAKCP-ENLLKNGACVAKCPK 282
QY 318 HNGEVTAEGRGTCRCKSKPCARVCYVGMGYIYANSKFIGITLEPAGCKIFGSLAPL 377
DB 283 GMPQNSE-----CVPCKGVCPKTCGEGE---VHSDNIG---NYKDTTIEGSLTL 329
QY 378 PESFDGDPASNT-----APLOPELOVFEETLEITGYLYISAMPDLSDLVQNDQ 429
DB 330 DQSFQFOCVQVYNTFSGFRPYIKIDPRLEVFSTVKEITGFINIGAHNPFTTLNYPRLNE 389

```

```

QY 430 VIRGRILHNGAY-STLLOGLISWLGRLSRLBLSGLALIHNNHLCFVHTVPMDOFRN 488
DB 390 VVGGROLKENLFASYIYIKTSIKSELKRVNVSGLVILENSDLCPVEDIDSEIKKS 449
QY 488 PHOALLHTANRPEDECVBGLAQHLCARGHCWGPPTQVCNCSQFLRGQCEVECEVLC 548
DB 450 SDHEVWVKRNATCHEHEGMECSQCSKACWGWGPGQCECNVYKXKCLDSCK--- 506
QY 549 GLPREY-VNARHCLCPHEPCORNGSVTCFPEADOCACAHYDPPRCFARGC----- 601
DB 507 SLPLYSVDSKTCGCHQCECKD----FCYGPNEBNCSCMNVKDRFCVAECPTTHAM 561
QY 602 -----SGVKPDLSTYMPWKFPD----- 618
DB 562 NGTCINCHKTCWCGRPDTIAPDGCISCDAIIGSDAKIERCLMKDESDCPDGYSDYVL 621
QY 619 -BEG----- 621
DB 622 QEEGPLKOLSGKAVRKCHPRCKCTGYGFHEPCQECTGYKGEQCEDECPDQFYANEE 681
QY 622 --ACQPCPINCT-----HSCVDL-----DD-----KCGPABQ----- 646
DB 682 TRICLPCHQECRGCHGLGDHHECRNLKLPEDGPYDNATFTCVSNCPASHPKRPOEA 741
QY 647 -----RASPLTISAVVGLLVVLCVVPGL---LIERQCKIRKRYTM 687
DB 742 GKIPGYCSADMSQGLRIEPOQVYKWSMALILLCVFGIAFVLFSSHKKKXAVVM 801
QY 688 RRLLOETELVEPLFSGAMPQAOARILKEGYIANSKFIIGTEL-TVYGGIMIPGENV 746
DB 802 TMLAGCEDSEPLRSNVGNPLTKLRIIKEKEIRGG-VLGMGAFGVFGVMMPGESV 860
QY 747 KIPVAIKVLRNTEBPKANKELLDEAYVAVGSPYVSRLLGICLTSTVQVLVTOIMPYGT 806
DB 861 KIPVAIKVLMEMSGESSEKFELEAYIMASVEHENLKLAVCMTSQMLITOLMLGCI 920
QY 807 LDHVENRGRGLSDILNMCQIAKMSYLEDEVLYVRDLAANVLYKSNHYKIDFGL 866
DB 921 LDYVRNKKDKTIGSKALLWSTQIARQAVYLEERLVHRDLAARVAVQTPSCVKITVFGI 980
QY 867 ARLLDIDETEVHADOGKVPKIMMALESILRRFTHQSDVMSYGVTWELMTFGAKPYDG 926
DB 981 AKLLDFDSEYRAAGKMPKIMALECIRHRVFPSKSDVAFGITIMELLTYGARYENV 1040
QY 927 PAREIPDLLEKGEELPQPTICTIVYKIMYKCMIDSECPRRRELVSFPRMADPQRF 986
DB 1041 PAKOVPELIEIGHKLPQPDICSLDVYCIILSCWVLADARPTFKOLAETFAEKABDPGRV 1100
QY 987 VVIQNEDELGPASPLDSTFYRSILEDMDMDLV----- 1018
DB 1101 LMI-----PQDKRMRLBPSYTNODEKDLIRTLAVYMAAAAAAGASVNDVPSTI 1151
QY 1019 -DAEEYLVPOQGFPCPDPAAGAGMVAHRRSSSTRSGGDLTLGLEPSEBEAPRS--- 1073
DB 1152 AETDEYLOPKTRPSIMLGPBA-----VEPS-DEMPKSLRYC 1187
QY 1074 --PLAF--SEGAASPVFDGDLGWAAGKGLQSLPTHDPSPLCQYSSDPVPLPSETDGYV 1128
DB 1188 KDPILKPDDETIDGHKEV-----GVGGIR-----LMLPLDDEDVL 1221
QY 1129 APLTCSPPQPEYVNPQVPRPSPSREGPLPAPBPAGATLERAKTLSPGKXGVKVFARFG 1188
DB 1222 MF-TCOSQ--NOS-----TPG-----YMDLIGVP 1242
QY 1189 GAVENPEYL-----TPQGAAPQHPBPAPAFSPAFDNLVYWDQDPFERGAPSPSTFG 1239
DB 1243 ASVNDPEYLMOSTQATAGLAQSGWG-FHTPP-----PPTWNG 1279
QY 1240 TPTAENPE 1247
DB 1280 MPTHQHSQ 1287

```

```

RESULT 8
Q9UK79 ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Herectatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).
RP [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF17761; AAD56009.2;
DR InterPro; IPR000494; EGFR_L domain.
DR Pfam; PF00757; Furin-like; 1.
DR SMART; PF01030; Recep_L domain; 1.
DR Pfam; SMO0261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEEIIE347E2D030C CRC64;

Query Match 27.5%; Score 1871; DB 4; Length 419;
Best Local Similarity 98.8%; Pred. No. 6.6e-134;
Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MELALCRWGLLALALPPGAASTQVCTGTDMLRLPASPETHDMLRHLVYQSGQVQGNL 60
DB 1 MELALCRWGLLALALPPGAASTQVCTGTDMLRLPASPETHDMLRHLVYQSGQVQGNL 60
QY 61 ELYTLPNNAISLFLQDIQEVQGYVLAHNOVROYVQRLRYRGTOLEFDNALAVLDNG 120
DB 61 ELYTLPNNAISLFLQDIQEVQGYVLAHNOVROYVQRLRYRGTOLEFDNALAVLDNG 120
QY 121 DPLNNTTPVYGAAPGSGRLREQLSLTEILKGYLIQRPQLCYQDITILKDIFFHXNQLA 180
DB 121 DPLNNTTPVYGAAPGSGRLREQLSLTEILKGYLIQRPQLCYQDITILKDIFFHXNQLA 180
QY 181 LTLIDTNRSPACHPSCMCKSGRCWGESSESDQSLRTVYCAAGCARCKGPLPTDCHEQC 240
DB 181 LTLIDTNRSPACHPSCMCKSGRCWGESSESDQSLRTVYCAAGCARCKGPLPTDCHEQC 240
QY 241 AAGGTGPKHSDCLACHFNHNSGICELHCPALVYVNTDRESMNPBGRITTFGASCTYACP 300
DB 241 AAGGTGPKHSDCLACHFNHNSGICELHCPALVYVNTDRESMNPBGRITTFGASCTYACP 300
QY 301 YNYLSTDVSGCTLVCPLNQEVYAEADGTORCEKCSKPCARVCYGL 345
DB 301 YNYLSTDVSGCTLVCPLNQEVYAEADGTORCEKCSKPCARVCYGL 345

RESULT 9
Q8R2X1 ID Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

```

RN [1]
 RP SEQUENCE FROM N.A.
 RA Scienceberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC027080; AAH27080.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 367 AA; 40163 MW; 0B503395F9E101B0 CRC64;

 Query Match 25.5%; Score 1739; DB 11; Length 367;
 Best Local Similarity 88.0%; Pred. No. 6e-124;
 Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

 QY 889 MALESILRRRFTGSDVMSYGVTVWELMTFGAKPYDGIIPAREIDPLEKEGRLPQPICT 948
 DB 1 MALESILRRRFTGSDVMSYGVTVWELMTFGAKPYDGIIPAREIDPLEKEGRLPQPICT 60
 QY 949 IDVYIMVYKCMWIDSECRPRRELVSSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 1008
 DB 61 IDVYIMVYKCMWIDSECRPRRELVSSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 120
 QY 1009 LEDDMGDLVAEEVLVPOQGFPCPDPAAGAGVHHRHSSSTRSGGDLTLGLEPSEE 1068
 DB 121 LEDDMGDLVAEEVLVPOQGFPCPDPAAGAGVHHRHSSSTRSGGDLTLGLEPSEE 180
 QY 1069 EAPRSPLAPSGAGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYV 1128
 DB 181 EAPRSPLAPSGAGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYV 240
 QY 1129 APLTCSQPEEVYNQPDVRFQPPSPREGPLPAAPAGATLERAKTLPFGKXGVKDVAFAG 1188
 DB 241 APLTCSQPEEVYNQPDVRFQPPSPREGPLPAAPAGATLERAKTLPFGKXGVKDVAFAG 300
 QY 1189 GAVENPEYLTPOGGAAPQPPSPAPFDNLVYMDPPERGAPEPTFGCTTAENPEY 1248
 DB 301 GAVENPEYLTPOGGAAPQPPSPAPFDNLVYMDPPERGAPEPTFGCTTAENPEY 360
 QY 1249 LGLDVPLV 1255
 DB 361 LGLDVPLV 367

 RESULT 10
 Q8WYV0 PRELIMINARY; PRT; 412 AA.
 AC Q8WYV0;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 44.7 kDa protein.
 GN P3659.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Primates; Catarrhini; Homnidae; Homo.
 CX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
 RA Wan D.F., Gu J.R.;
 RT "Novel human cDNA clones with function of inhibiting cancer cell
 RT growth."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF18349; AAL5586.1; -
 DR InterPro; IPR002048; Ee-hand.
 DR InterPro; IPR000719; Euk_kinase.
 DR InterPro; IPR001245; Tyr_kinase.
 DR InterPro; IPR004019; YLP motif.
 DR Pfam; PF00069; Kinase; 1.
 DR Pfam; PF02757; YLP; 2.
 DR ProDom; PD000001; Euk_kinase; 1.
 DR SMART; SM00219; TyKc; 1.
 DR PROSITE; PS00018; EF_HAND; UNKNOWN; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KW Hypothetical protein.

SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F72D2BC CRC64;

 Query Match 24.9%; Score 1697.5; DB 4; Length 412;
 Best Local Similarity 80.5%; Pred. No. 1e-120;
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

 QY 889 MALESILRRRFTGSDVMSYGVTVWELMTFGAKPYDGIIPAREIDPLEKEGRLPQPICT 948
 DB 1 MALESILRRRFTGSDVMSYGVTVWELMTFGAKPYDGIIPAREIDPLEKEGRLPQPICT 60
 QY 949 IDVYIMVYKCMWIDSECRPRRELVSSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 1008
 DB 61 IDVYIMVYKCMWIDSECRPRRELVSSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 120
 QY 1009 LEDDMGDLVAEEVLVPOQGFPCPDPAAGAGVHHRHSSSTRSGGDLTLGLEPSEE 1068
 DB 121 LEDDMGDLVAEEVLVPOQGFPCPDPAAGAGVHHRHSSSTRSGGDLTLGLEPSEE 180
 QY 1069 EAPRSPLAPSGAGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYV 1128
 DB 181 EAPRSPLAPSGAGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYV 240
 QY 1129 APLTCSQPEEVYNQPDVRFQPPSPREGPLPAAPAGATLERAKTLPFGKXGVKDVAFAG 1188
 DB 241 APLTCSQPEEVYNQPDVRFQPPSPREGPLPAAPAGATLERAKTLPFGKXGVKDVAFAG 300
 QY 1189 GAVENPEYLTPOGGAAPQPPSPAPFDNLVYMDPPERGAPEPTFGCTTAENPEY 1248
 DB 301 GAVENPEYLTPOGGAAPQPPSPAPFDNLVYMDPPERGAPEPTFGCTTAENPEY 360
 QY 1221 YVMDPPER-----GAPSTFGCTTAEN 1245
 DB 361 YVMDPPER-----GAPSTFGCTTAEN 410

 RESULT 11
 Q86712 PRELIMINARY; PRT; 729 AA.
 AC Q86712;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Polypeptide.
 GN POLYPEPTIDE.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
 CX NCBI_Taxid=11950;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; PubMed=8152791;
 RA Vennstrom B., Raynosccheck C., Jansson L., Doederlein G., Lhotak V.,
 RA Johnson A., Beng H.;
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities."
 RL Oncogene 9:1307-1320(1994).
 DR EMBL; S59372; AAC60725.1; -
 DR HSP; P03322; 1A6S.
 DR InterPro; IPR000719; Euk_kinase.
 DR InterPro; IPR004028; Retro M.
 DR InterPro; IPR001245; Tyr_kinase.
 DR Pfam; PF00069; Kinase; 1.
 DR Pfam; PF02813; Retro M; 1.
 DR ProDom; PD000001; Euk_kinase; 1.
 DR SMART; SM00219; TyKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 729 AA; 80649 MW; 84DF6514FE1D63 CRC64;

 Query Match 24.1%; Score 1640; DB 15; Length 729;
 Best Local Similarity 53.2%; Pred. No. 5.7e-116;

Matches 348; Conservative 75; Mismatches 127; Indels 104; Gaps 17;

QY 569 PONGSVTCGPEADOCVACAHYKDPFCVACPSGVKNDLSYMIWKPPEBEGCCPCPI 628
 DB 141 PEETATPKTGP--DHCKCAHFIDGPPCVACCPAGVIGENDTL-VKRYADANAACQOLCHP 197
 QY 629 NCTHSCVDLDKCPACPAEORASPLTISAVV-GILLVVLGVVFGILIKRQCKIRKRYTM 687
 DB 198 NCTRGCKGFGLEGCP---NGSKTSPSIAGVVGGLCLVWGLGIGLYLRR-HIVRRTL 253

QY 688 RLLOETELVPLTSSGAMPNACMRILKEQYIKANSKFIGITEL-TVYKGIWIPGENV 746
 DB 254 RLLOERELVPLTSSGAMPNACMRILKEQYIKANSKFIGITEL-TVYKGIWIPGENV 312

QY 747 KIPVAIKVLRNTSPKANKELIDEAVYVAGVSPVYSRLIGICTSTVOLTPMYPGCL 806
 DB 313 KIPVAIKVLRNTSPKANKELIDEAVYVAGVSPVYSRLIGICTSTVOLTPMYPGCL 372

QY 807 LDHVRNRRGLSGDILWKCQIAGKMSYLEVDVLRHDLAARNVLKSPNHVKITDFGL 866
 DB 373 LDYIREHKNISQYLLWVCQIAGKMSYLEVDVLRHDLAARNVLKSPNHVKITDFGL 432

QY 867 ARLLDIDETEHADGKVPDIKMALESILRRRPTHOSDWSYGVTVWELMTFGAKPYDGI 926
 DB 433 AKLLGADKEVYHABEGKVPDIKMALESILRRRPTHOSDWSYGVTVWELMTFGAKPYDGI 492

QY 927 PAREIPDLLEKGERLPDPPICTIDVYIMVWCMIDSECRPRELVSFESRMARDPQR 986
 DB 493 PAREIPDLLEKGERLPDPPICTIDVYIMVWCMIDSECRPRELVSFESRMARDPQR 552

QY 987 VVIO-NEDLGPASPLDSTFYRSLLEDMDGLVAEYLVPOGFCPPRAGAGMTHH 1045
 DB 553 LVIOGGERMHLPSPTDSKFRYRTLMSEEDMEDIVDAEYLVPHQGF 598

QY 1046 RHRSSTRSGGDLTLGLEPSEEBAPRSP-----APSEGASDVDFDGLMGAKGLQS 1100
 DB 599 --NSPST-----SRTPLSLSLATSNNSATNCID-----RNGQG 630

QY 1101 LPTHDPSPLOQRYSEPTVPLPSET--DGVAAPLTCSPOPEYVNOQPDVPRPSPRESGFLP 1158
 DB 631 HVRREDSFVQRYSSPTNPFLEESIDGFL-----PAPEYVNO--LMPKPS----- 675

QY 1159 AARPAGATLERAKTLPCKNGVVKDF-----AFGAVENPEYL 1197
 DB 676 -----TAMVQNIYNNISLTAISKLPMSDRYONSHSTAVDNPETL 715

RESULT 12
 Q86714 PRELIMINARY; PRT; 567 AA.

AC 086714;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE V-erbB protein (Fragment).
 GN V-ERBB.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
 OC NCBI_TaxID=11950;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; PubMed=8152791;
 RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
 RA Johnson A., Beug H.;
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities.";
 RL Oncogene 9:1307-1320(1994).
 DR EMBL; S69372; AAC60727.1; .
 DR HSP; F11362; IFGK.
 DR InterPro; IPR000719; Euk_kinase.
 DR InterPro; IPR001245; Tyr_kinase.
 DR Pfam; PF00069; kinase; 1.

DR PRINTS; P800109; TYRKINASE.
 DR ProDom; P8000001; Euk_kinase; 1.
 DR SMART; SM00219; TyrcK; 1.
 DR PROSITE; P800107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; P850011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; P800109; PROTEIN_KINASE_TYR; 1.
 FT NON TER
 FT 1
 SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF25EL CRC64;

Query Match 24.0%; Score 1638; DB 15; Length 567;
 Best Local Similarity 53.8%; Pred. No. 5,56-116;
 Matches 347; Conservative 73; Mismatches 121; Indels 104; Gaps 17;

QY 578 GPEADQCVACAHYKDPFCVACPSGVKNDLSYMIWKPPEBEGCCPCPI 637
 DB 1 GP--DHCKCAHFIDGPPCVACCPAGVIGENDTL-VKRYADANAACQOLCHPCTRGCKGP 57

QY 638 DDKGPACPAEORASPLTISAVV-GILLVVLGVVFGILIKRQCKIRKRYTMRLLOETEL 696
 DB 58 GLEGCP--NSGKTPSIAGVVGGLCLVWGLGIGLYLRR-HIVRRTLRRLLOEREL 113

QY 697 VEPPLPSCGAMPNACMRILKEQYIKANSKFIGITEL-TVYKGIWIPGENVKIPVAIKVL 755
 DB 114 VEPPLPSCGAMPNACMRILKEQYIKANSKFIGITEL-TVYKGIWIPGENVKIPVAIKVL 172

QY 756 RENTSPKANKELIDEAVYVAGVSPVYSRLIGICTSTVOLTPMYPGCLLDHVRNRRG 815
 DB 173 RENTSPKANKELIDEAVYVAGVSPVYSRLIGICTSTVOLTPMYPGCLLDHVRNRRG 232

QY 816 RIGSDILWKCQIAGKMSYLEVDVLRHDLAARNVLKSPNHVKITDFGLARLLDIDET 875
 DB 233 NIGSQYLLWVCQIAGKMSYLEVDVLRHDLAARNVLKSPNHVKITDFGLARLLDIDET 292

QY 876 EYHAGGVKPIKMALESILRRRPTHOSDWSYGVTVWELMTFGAKPYDGI 935
 DB 293 EYHAGGVKPIKMALESILRRRPTHOSDWSYGVTVWELMTFGAKPYDGI 352

QY 936 EKGERLPDPPICTIDVYIMVWCMIDSECRPRELVSFESRMARDPQRVVIQ-NEDL 994
 DB 353 EKGERLPDPPICTIDVYIMVWCMIDSECRPRELVSFESRMARDPQRVVIQ-NEDL 412

QY 995 GPASPLDSTFYRSLLEDMDGLVAEYLVPOGFCPPRAGAGMTHHRRSSSTRS 1054
 DB 413 HLPSPDSKFRYRTLMSEEDMEDIVDAEYLVPHQGF-----NSPST-- 454

QY 1055 GGGDLTLGLEPSEEBAPRSP-----APSEGASDVDFDGLMGAKGLQSLPTHDPSP 1109
 DB 455 -----SRTPLSLSLATSNNSATNCID-----RNGGHRVREDSFV 490

QY 1110 QRYSDPTVPLPSET--DGVAAPLTCSPOPEYVNOQPDVPRPSPRESGFLP 1167
 DB 491 QRYSDPTVPLPSET--DGVAAPLTCSPOPEYVNOQPDVPRPSPRESGFLP 526

QY 1168 ERAKTLPCKNGVVKDF-----AFGAVENPEYL 1197
 DB 527 -----TAMVQNIYNNISLTAISKLPMSDRYONSHSTAVDNPETL 566

RESULT 13
 Q64895 PRELIMINARY; PRT; 962 AA.

AC 064895;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Gag-V-erb-A, v-erb-B protein.
 GN GAG-V-ERB-A, V-ERB-B.
 OS Avian erythroblastosis virus.
 OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
 OC NCBI_TaxID=11861;
 RN [1]
 RP SEQUENCE FROM N.A.

```

RX MEDLINE:90206603; PubMed:1969616;
RA Brustin A., Jackson J., Bishop J.M., McCauley D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the
RT transforming potential of the oncogene v-erb-B."
RL Oncogene 5:15-24(1990).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY)
-1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL: X52209; CA36459.1; -.
DR HSSP: P10828; 2NLL.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR000536; Hormone_rec_119.
DR InterPro: IPR001723; Stehnm_receptor.
DR InterPro: IPR001245; Tyr_pkinase.
DR InterPro: IPR001628; Zn_f_Csteroi.
DR Pfam: PF00104; hormone_rec_1.
DR Pfam: PF00069; kinase_1.
DR PRINTS: PR00398; STRDHOMONR.
DR PRINTS: PR00047; STRDIPINGER.
DR PRODOM: PD000001; Euk_pkinase_1.
DR PRODOM: PD000035; Zn_f_Csteroi.
DR SMART: SM00430; HOL1; 1.
DR SMART: SM00219; Tyrc; 1.
DR SMART: SM00399; ZNF_C4; 1.
DR PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Transferase; Tyrosine-protein kinase;
KW Zinc-finger.
SQ
SEQUENCE 962 AA; 108320 MW; 3C5AD791EAE95CE CRC64;

Query Match 23.2%; Score 1581.5; DB 15; Length 962;
Best Local Similarity 50.4%; Pred. No. 2.4e-111;
Matches 349; Conservative 73; Mismatches 150; Indels 121; Gaps 20;

QY 541 VEEGRVLOGLPRE-VYNNR-HCLP-----CHPEQC 568
DB 354 IEKCESTYLLAFHEHINRKNNITHFMSKILMKVADLMIGAYHASFELMKVECPTELS 413
QY 569 PQNGSVTCFGPEADQCVACAHYKDPFCVACPSGVKXPDLSYMPMKFPDEGACQPCPI 628
DB 414 PQE-----VGP--DHCKCAHFIDGPHCVKACPAAGVLGENDTL-VMKYADANAVCQLCHP 465
QY 629 NCHTSCVDLDKGCAPBQASPLTSISAVV-GILLVVLGVVFGILLIKRQCKIRKRYM 687
DB 466 NCTRGCKGPGLEGCP--NGSKTPSIAAGVGGLLCLVWGLGIGLYLRRR-HIVKRTLL 521
QY 688 RRLQETELVEPLTPSGAMPNQAOIRLKEQYIKANSKFTIGTEL-TVYKGIWIPGGENV 746
DB 522 RRLQERELVEPLTPSGARPQALRLIKETEPK-KYKVLGFGAFTVYKGLMIPGGENV 580
QY 747 KIPVAIKYLRNTPSPKANKELLDEAYVAVAGVSPYSRLIGICTSTVQLVTOIMPYGC 806
DB 581 TIPVAIKELREATSPKANKELLDEAYVAVASVDNPHVCRLLGICTSTVQITQIMPYGCL 640
QY 807 LDHYRENGRGLSGODLLNMCQIAKMSYLEDVRLVHRDLAARNVLYKSPNNHYKITDFGL 866
DB 641 LDYIREHKDNIGSYLLNMCVQIAKGNVLYEERHMYRDLAARNVLYKTPQHNKIDTFGL 700
QY 867 ARLLIDETETEHADGGKVPKIKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI 926
DB 701 AKQLGADKEVHAEGKVPKIKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKRYDGI 760
QY 927 PAREIPDLLEKGEELPQPTCTITVYVIMTKCMWIDSECPRRRELVSFSSMARDPQGF 986
DB 761 PASEISSVLEKGEELPQPTCTITVYVIMTKCMWISGDSRPKRELIAEFSKXARDPPRY 820
QY 987 VVIG-NEDLGSPFLDSTFYRSLDEDDMDGLVDAEEYLVPOQGFCCPPAPGAGGMVNH 1045
DB 821 LVIGDERMHLPSPTDCKFYRTLMEEEDMEDIVADADYLVPHQGF----- 866

```

```

QY 1046 RHRSSTRGSGDUTLGLPEPEEAPRSPPLAPSEGAGSDVFDGLGMAAGLQSLPTH 1105
DB 867 --NSPST-----STRPLISSLSASN-----NSATKICLRNGH- 898
QY 1106 PSPFLORYSEDTPLPSTGDIYAPLTCSPQPEYVNOPDVRRPQSPSEEGPLPAPAGA 1165
DB 899 -----PVREDFGL-----PAPEYVNO--LMPKXSTAWQNIYNYISL 935
QY 1166 T-LEBAKTLSPKNGVVDVPAFGAVENPEYL 1197
DB 936 TAISKLPMDSRYN-----SHSTAVDNPEYL 961

```

RESULT 14

Q85468

ID Q85468

PRELIMINARY; PRT; 545 AA.

NC Q85468

01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Avian Erythroblastosis virus (T334) v-erbB gene.

OS Avian erythroblastosis virus.

OC Viruses; Retroviruses; Retroviridae; Avian type C retroviruses.

OX NCBI_Taxid=11861;

RN (1)

RP SEQUENCE FROM N. A.

RX MEDLINE=88217326; PubMed=2897102;

RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Hayman M.J.;

RT "Common site of mutation in the erbB gene of avian erythroblastosis

RT virus mutants that are temperature sensitive for transformation.";

RL Oncogene Res. 1:265-278(1987).

DR EMBL: X06943; CA30024.1; -.

DR HSSP: P11362; 1FEK.

DR InterPro: IPR000719; Euk_pkinase.

DR InterPro: IPR001245; Tyr_pkinase.

DR Pfam: PF00069; kinase_1.

DR PRODOM: PD000001; Euk_pkinase_1.

DR SMART: SM00219; Tyrc; 1.

DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.

DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.

KW ATP-binding; Transferase; Tyrosine-protein kinase.

SQ SEQUENCE 545 AA; 60899 MW; 140DCE8CCAD0F8A74 CRC64;

```

Query Match 23.1%; Score 1573; DB 15; Length 545;
Best Local Similarity 53.4%; Pred. No. 4.6e-111;
Matches 336; Conservative 70; Mismatches 129; Indels 94; Gaps 17;

QY 578 GPEADQCVACAHYKDPFCVACPSGVKXPDLSYMPMKFPDEBACQPCPICTHSCVDL 637
DB 1 GP--DHCKCAHFIDGPHCVKACPAAGVLGENDTL-VMKYADANAVCQLCHPCTRGCKGP 57
QY 638 DDKGCPAQRASPLTSISAVV-GILLVVLGVVFGILLIKRQCKIRKRYMRLQETEL 696
DB 58 GLEGCP--NSKTPSIAAGVGGLLCLVWGLGIGLYLRR-HIVKRTLLRLDQREL 113
QY 697 VEPITPSGAMPNQAOIRLKEQYIKANSKFTIGTEL-TVYKGIWIPGGENVXKIPVAIKYL 755
DB 114 VEPITPSGAMPNQAOIRLRIKETEPR-KYKVLGFGAFTVYKGLMIPGGENVTLIPVAIKEL 172
QY 756 RENTSPKANKELLDEAYVAVAGVSPYSRLIGICTSTVQLVTOIMPYGLDHYRENG 815
DB 173 RENTSPKANKELLDEAYVAVASVDNPHVCRLLGICTSTVQLVTOIMPYGLDHYREHND 232
QY 816 RLGSQDLLNMCQIAKMSYLEDVRLVHRDLAARNVLYKSPNNHYKITDFGLARLLIDET 875
DB 233 NIGSYLLNMCVQIAKGNVLYEERHMYRDLAARNVLYKTPQHNKIDTFGLAKQLGADK 292
QY 876 EYHADGGKVPKIKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDL 935
DB 293 EYHAEKGKVPKIKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKRYDGIIPASEISSVL 352

```

```

QY 936 EKERLPPPICTIDVYIMVCMWIDSECRPREFELSEFSRMARDQREVIQ-NEDL 994
DB 353 EKERLPPPICTIDVYIMVCMWIDSECRPREFELSEFSRMARDQREVIQ-NEDL 412
QY 995 GPASPLDSTFFYSLIEDDDMGDLVAEEYLVQCGFPCPDPAAGVAGVHHRRSSSTRS 1054
DB 413 HLPSPDTSKFRYRLMEEDMEDIVIDADELYPHQOFF-----NSPST-- 454
QY 1055 GGGDLTLGLEPSEEAERSP-----APSEGAAGSVFQGDLCMGAKGLQSLPTHPDSPL 1109
DB 455 -----SRTPLLSSLSATSNNSATNCTIRBNG-----H----- 481
QY 1110 QRYSEDPVPLPSEIDGYVAPLTCSPQPEYVQPDVRQPPSPREGPLPAAPAGAT-LE 1168
DB 482 -----PYREDGFL-----PAPEYVQ--LMPKPSSTAMVQ-QIYNYISLTAIS 522
QY 1169 RAKTSPGKGVGVKQVAFAGAVENPEYL 1197
DB 523 KLPMDSRVQ-----SHSTAVDNPEYL 544

RESULT 15
Q9WVF5 PRELIMINARY; PRT: 655 AA.
AC Q9WVF5;
DT 01-NOV-1999 (TEMBLrel. 12, Created)
DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
DE 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielsen A.J., Scheil C.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maithe N.J.;
RT "Alternative transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVYAC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maithe N.J.;
RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egfr transcripts encoding truncated receptor isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai U., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Akizawa T., Hara A., Fukunishi Y., Kono H., Aachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleschmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudt P., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Oikido T., Furuno M., Aono H., Baldarelli R., Barin G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima M., Mazzarelli J., Mommaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

```

```

RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,
RA Wymbs-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohzuki S.,
RA Hayaishiaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AF124513; AAD4149.1; -.
DR EMBL; AF275366; AAG28047.1; -.
DR EMBL; AF275366; AAG28047.1; JOINED.
DR EMBL; AF275366; AAG28047.1; JOINED.
DR EMBL; AK004944; BAB23688.1; -.
DR EMBL; AK004883; BAB23641.1; -.
DR EMBL; AK004911; BAB23662.1; -.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; Fu; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

```

```

Query Match      22.1%; Score 1506.5; DB 11; Length 655;
Best Local Similarity 44.5%; Pred. No. 7e-106;
Matches 285; Conservative 97; Mismatches 233; Indels 25; Gaps 9;

```

```

QY 11 LTLALLPGAA--STVCTGDMKRLRASPETHLDMRLHYOCQVQGNLELTYPTN 68
DB 14 LTLALCAAGALAEKVKQGSNNLTGCTGDEHFLSLGRYNNNECVLLNLETTYQRN 73
QY 69 ASLSFLQDIQYQGVYLIHNVQVRLRIRYRGTQLPEDNYALALVDNGDEPLNTP 128
DB 74 YLSLFLKTIQEVAGVLLALNTVRIRPLENQIIRGNALYENTYALALISN----- 124
QY 129 VTGASPGGRLRELRLSTLEIKGGLVQLQNPQLCYOPTLMKDI-----FKNNQLATLI 184
DB 125 -YGRTRGRLRELPENMLDEILIGAVRSNPILNMOTIQRDIVQVNFNSMSMDL--- 180
QY 185 DTRSRACHPSPCMKSGRCWGESSEDDQSLTRTVCAAGCA-RCKGFLPTDCCHEQCAAG 243
DB 181 -QSHPSCPKCDPSCPNQSCWGGGEENQKTKIKIAQCCSHRCGRSPSCGHNQCAAG 239
QY 244 CTGPRGSCCLACIHNHSGIEHLCPALVYNTTFESMRPERRYTFGASCYTAGYNY 303
DB 240 CTGPRSCDLVCQFQDEATKOTCPPLMLNPTTYQMDVNPBGKVSFGATCVKCKPRNY 299
QY 304 ISTDVGSCTLVCPILHNOEVTAEQTCRCEKSKPCARVCYGLGQVYKASKFITELE 363
DB 300 VVTDHGSQVRAQGPVYEV-BEDGIRKCKKDCGRCVNCIGIGERK-DTSLNATNIK 357
QY 364 -FAGCKKTFGSLAFIPSEFDGDPASNTAPLOPELOQVFTLEETGYLYISAMPDSLPL 422
DB 358 HFXYCTALSGDLHLPLVAFKDSFTTRPPDPRLEILKTKVETGFLLIQAMPDMNTDL 417
QY 423 SVFOVLQVIRGRILHNGAVSLTGLGLGSIWGLSRLSGSLALIHNNHLPFVHTVP 482
DB 418 HAFENLEIRGRTRQKHQGFSLAVGLNITSGLRSLEISDGVITISGNRLCTANTIM 477
QY 483 DQLFNPHQALLHTAMREDECVGEGLACHQLCARGHGCMWGPPTQVNCISQPLRGQEV 542
DB 478 KTLFETPQKTKIMNPRAEKCKKLVNHVNCPLGSESCWGPPEPRDVCQNVARGRECV 537
QY 543 ECRVYQGLPREYVNAHRLCPHRECCQONQSNVYCFGEADQCVACAHYKRPVCVARGPS 602
DB 538 KNITLGGAPREFVENSEIQCHPCLDQANNITGTGSPGNCTQCAHYIDGPHCVATCPA 597
QY 603 GVKPDLSTYPIWKEPDEGACQCPINCTHSQVLDLDRGC 642
DB 598 GIMGENNTL-VWXTAADANNVCHLCHACTGCGAGPGIGGC 636

```

Search completed: July 22, 2003, 09:24:34
 Job time: 51.8246 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:41:54 ; Search time 36.9339 Seconds

(without alignments)

4527.811 Million cell updates/sec

Title: SEQ4-695-709-12

Perfect score: 6815

Sequence: 1 MEALALCRWGLLALLPPGA.....TFKGTPTANPEYLGDVPR 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6646	97.5	1255	21	AA192620 Human heregulin 2
2	6646	97.5	1255	22	AA12130 Human tyrosine kin
3	6646	97.5	1255	22	AA12130 Human tyrosine kin
4	6646	97.5	1255	23	AA12130 Human tyrosine kin
5	6646	97.4	1255	17	AA12130 Human tyrosine kin
6	6640	97.4	1255	20	AA12130 Human tyrosine kin
7	6640	97.4	1255	21	AA12130 Human tyrosine kin
8	6640	97.4	1255	21	AA12130 Human tyrosine kin
9	6640	97.4	1255	22	AA12130 Human tyrosine kin
10	6640	97.4	1255	22	AA12130 Human tyrosine kin

11	6640	97.4	1255	23	AA12130 Human tyrosine kin
12	6640	97.4	1255	23	AA12130 Human tyrosine kin
13	6640	97.4	1255	23	AA12130 Human tyrosine kin
14	6640	97.4	1255	23	AA12130 Human tyrosine kin
15	6640	97.4	1255	23	AA12130 Human tyrosine kin
16	6640	97.4	1255	23	AA12130 Human tyrosine kin
17	6640	97.4	1255	23	AA12130 Human tyrosine kin
18	6640	97.4	1255	23	AA12130 Human tyrosine kin
19	6640	97.4	1255	23	AA12130 Human tyrosine kin
20	6640	97.4	1255	23	AA12130 Human tyrosine kin
21	6640	97.4	1255	23	AA12130 Human tyrosine kin
22	6640	97.4	1255	23	AA12130 Human tyrosine kin
23	6640	97.4	1255	23	AA12130 Human tyrosine kin
24	6640	97.4	1255	23	AA12130 Human tyrosine kin
25	6640	97.4	1255	23	AA12130 Human tyrosine kin
26	6640	97.4	1255	23	AA12130 Human tyrosine kin
27	6640	97.4	1255	23	AA12130 Human tyrosine kin
28	6640	97.4	1255	23	AA12130 Human tyrosine kin
29	6640	97.4	1255	23	AA12130 Human tyrosine kin
30	6640	97.4	1255	23	AA12130 Human tyrosine kin
31	6640	97.4	1255	23	AA12130 Human tyrosine kin
32	6640	97.4	1255	23	AA12130 Human tyrosine kin
33	6640	97.4	1255	23	AA12130 Human tyrosine kin
34	6640	97.4	1255	23	AA12130 Human tyrosine kin
35	6640	97.4	1255	23	AA12130 Human tyrosine kin
36	6640	97.4	1255	23	AA12130 Human tyrosine kin
37	6640	97.4	1255	23	AA12130 Human tyrosine kin
38	6640	97.4	1255	23	AA12130 Human tyrosine kin
39	6640	97.4	1255	23	AA12130 Human tyrosine kin
40	6640	97.4	1255	23	AA12130 Human tyrosine kin
41	6640	97.4	1255	23	AA12130 Human tyrosine kin
42	6640	97.4	1255	23	AA12130 Human tyrosine kin
43	6640	97.4	1255	23	AA12130 Human tyrosine kin
44	6640	97.4	1255	23	AA12130 Human tyrosine kin
45	6640	97.4	1255	23	AA12130 Human tyrosine kin

ALIGNMENTS

RESULT 1	AA192620	standard; Protein: 1255 AA.
ID	AA192620	
AC	AA192620	
XX	10-AUG-2000	(first entry)
XX	Human heregulin 2 (Her2)	
DE	Human heregulin 2 (Her2)	
XX	Her2: Vaccination; cytotoxic T-lymphocyte immunity;	
KW	self-protein; cancer; breast cancer; prostate cancer;	
KW	cell-associated peptide antigen; foreign epitope.	
OS	Homo sapiens.	
XX		
PH	Key	Location/Qualifiers
FT	Domain	1..173
FT		/label= "N-terminal"
FT		/note= "mature polypeptide"
FT	Region	5..25
FT		/label= "insertion region"
FT		/note= "suitable for foreign epitope insertion"
FT	Region	59..73
FT		/label= "insertion region"
FT		/note= "suitable for foreign epitope insertion"
FT	Region	103..117
FT		/label= "insertion region"
FT		/note= "suitable for foreign epitope insertion"
FT	Region	149..163
FT		/label= "insertion region"
FT		/note= "suitable for foreign epitope insertion"
FT	Domain	174..323

FT /label= Cysteine_rich_domain
 FT Region 210..224
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 250..264
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain 324..483
 FT /label= ligand_binding_domain
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 359..383
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 455..479
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain 484..623
 FT /label= Cysteine_rich_domain
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 579..593
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain 624..654
 FT /label= Transmembrane_domain
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 632..652
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 653..667
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain 655..1010
 FT /label= Tyrosine_kinase_domain
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 661..675
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 695..709
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 710..730
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain 1011..1235
 FT /label= C-terminal_domain

XX MO200020027-A2.
 XX 13-APR-2000.
 XX 05-OCT-1999; 99MO-DK00525.
 XX 05-OCT-1998; 98DX-0001261.
 PR 20-OCT-1998; 98US-0105011.
 XX
 PA (MEBI-) M & E BIOTECH AS.
 XX
 PI Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
 PI Gautam A, Birk P, Karlsson G;
 XX
 DR WPI, 2000-349917/30.
 DR N-PSDB; AAA09455.
 XX
 PT Inducing immune responses to weakly immunogenic, tumor associated
 PT peptide antigens for the treatment of breast and prostate cancer
 XX
 PS Claim 62; Page 193-198; 220PP; English.
 XX
 CC This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 CC Her2 can be used in the claimed method as an autovaccine to induce a CTL
 CC response. Subdominant CTL epitopes, antibody binding regions and
 CC cysteine residues involved in disulfide bonds are preserved in the
 CC immunogenized forms. Regions suitable for the insertion of foreign T
 CC helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the animal immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respective
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/breast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.

XX Sequence 1255 AA;

Query Match 97.5%; Score 6646; DB 21; Length 1255;
 Best Local Similarity 97.7%; Pred. No. 0;
 Matches 1227; Conservative 7; Mismatches 20; Indels 2; Gaps 2;

QY 1 MELALICRMGLLALLPPGAASVCTGTDMKRLRPASPTHLDMLRYOGQVQGNL 60
 DB 1 MELALICRMGLLALLPPGAASVCTGTDMKRLRPASPTHLDMLRYOGQVQGNL 60
 QY 61 ELTYLPTNASLFLDDIOEVQGVYLIANOVQVPLQRLIRVGTQLEEDNYALAVDNG 120
 DB 61 ELTYLPTNASLFLDDIOEVQGVYLIANOVQVPLQRLIRVGTQLEEDNYALAVDNG 120
 QY 121 DPLNNTPTPTGASPGGLREQLQRLSTELTKGVYLIQRPOLCYQDTIMKQIFPKNOLA 180
 DB 121 DPLNNTPTPTGASPGGLREQLQRLSTELTKGVYLIQRPOLCYQDTIMKQIFPKNOLA 180
 QY 122 DPLNNTPTPTGASPGGLREQLQRLSTELTKGVYLIQRPOLCYQDTIMKQIFPKNOLA 180
 DB 122 DPLNNTPTPTGASPGGLREQLQRLSTELTKGVYLIQRPOLCYQDTIMKQIFPKNOLA 180
 QY 181 LTLIDTNRBRACHPCSPMKSGRCWGESSEDDQSILTRVACGACGACGRLPTDCHEQC 240
 DB 181 LTLIDTNRBRACHPCSPMKSGRCWGESSEDDQSILTRVACGACGACGRLPTDCHEQC 240
 QY 181 LTLIDTNRBRACHPCSPMKSGRCWGESSEDDQSILTRVACGACGACGRLPTDCHEQC 240
 DB 181 LTLIDTNRBRACHPCSPMKSGRCWGESSEDDQSILTRVACGACGACGRLPTDCHEQC 240
 QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNDEGYTTGASCVTACP 300
 DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNDEGYTTGASCVTACP 300
 QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNDEGYTTGASCVTACP 300
 DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNDEGYTTGASCVTACP 300
 QY 301 YNVLSTDVGSCTLVCPLNQVEVTAEDGTQRCCKSPCARCYGIGMGIYIKANSKFIGIT 360
 DB 301 YNVLSTDVGSCTLVCPLNQVEVTAEDGTQRCCKSPCARCYGIGMGIYIKANSKFIGIT 360
 QY 301 YNVLSTDVGSCTLVCPLNQVEVTAEDGTQRCCKSPCARCYGIGMGIYIKANSKFIGIT 360
 DB 301 YNVLSTDVGSCTLVCPLNQVEVTAEDGTQRCCKSPCARCYGIGMGIYIKANSKFIGIT 360
 QY 361 ELFPACCKTIFGSLAFLEPSFDGDSANTAPLQBPQLQVFTELEITGYIISAMPDGLP 420
 DB 361 ELFPACCKTIFGSLAFLEPSFDGDSANTAPLQBPQLQVFTELEITGYIISAMPDGLP 420
 QY 361 ELFPACCKTIFGSLAFLEPSFDGDSANTAPLQBPQLQVFTELEITGYIISAMPDGLP 420
 DB 361 ELFPACCKTIFGSLAFLEPSFDGDSANTAPLQBPQLQVFTELEITGYIISAMPDGLP 420
 QY 421 DLSVFQNLQVIRGLIHNGAYSILTIQGISWLGISRLRELSGLALIHNTHLCEFTV 480
 DB 421 DLSVFQNLQVIRGLIHNGAYSILTIQGISWLGISRLRELSGLALIHNTHLCEFTV 480
 QY 421 DLSVFQNLQVIRGLIHNGAYSILTIQGISWLGISRLRELSGLALIHNTHLCEFTV 480
 DB 421 DLSVFQNLQVIRGLIHNGAYSILTIQGISWLGISRLRELSGLALIHNTHLCEFTV 480
 QY 481 PMDQLFRNPHQALLHTANRDEDECEVGEGLACHQLCARGHCWGPGPTQCVCNSQFLRQEC 540
 DB 481 PMDQLFRNPHQALLHTANRDEDECEVGEGLACHQLCARGHCWGPGPTQCVCNSQFLRQEC 540
 QY 481 PMDQLFRNPHQALLHTANRDEDECEVGEGLACHQLCARGHCWGPGPTQCVCNSQFLRQEC 540
 DB 481 PMDQLFRNPHQALLHTANRDEDECEVGEGLACHQLCARGHCWGPGPTQCVCNSQFLRQEC 540
 QY 541 VEECRVLOGLPREYVNAARHCLPQHECCOPQNSVTCFGEADQCYACAHYDPPCVARC 600
 DB 541 VEECRVLOGLPREYVNAARHCLPQHECCOPQNSVTCFGEADQCYACAHYDPPCVARC 600
 QY 541 VEECRVLOGLPREYVNAARHCLPQHECCOPQNSVTCFGEADQCYACAHYDPPCVARC 600
 DB 541 VEECRVLOGLPREYVNAARHCLPQHECCOPQNSVTCFGEADQCYACAHYDPPCVARC 600
 QY 601 PSGVRPDLSTYMPIMKRPDEGACQCPINCTHSCVYDLDDKGPAPORASPLTSYSAVVG 660
 DB 601 PSGVRPDLSTYMPIMKRPDEGACQCPINCTHSCVYDLDDKGPAPORASPLTSYSAVVG 660
 QY 601 PSGVRPDLSTYMPIMKRPDEGACQCPINCTHSCVYDLDDKGPAPORASPLTSYSAVVG 660
 DB 601 PSGVRPDLSTYMPIMKRPDEGACQCPINCTHSCVYDLDDKGPAPORASPLTSYSAVVG 660
 QY 661 ILLVVLGVVFGILIKRQOKIRKXTMRLLQETELVPLTPSGAMPNQAOMRLKEQYI 720
 DB 661 ILLVVLGVVFGILIKRQOKIRKXTMRLLQETELVPLTPSGAMPNQAOMRLKEQYI 720
 QY 661 ILLVVLGVVFGILIKRQOKIRKXTMRLLQETELVPLTPSGAMPNQAOMRLKEQYI 720
 DB 661 ILLVVLGVVFGILIKRQOKIRKXTMRLLQETELVPLTPSGAMPNQAOMRLKEQYI 720
 QY 721 KNSKRIGITEL-TYKKGWIDGDNWIKIPVAKYKLENTSPKAKKELLDAYVMAAGVGS 779
 DB 721 R-KVKVLGSGAGFTYKGIWIDGDNWIKIPVAKYKLENTSPKAKKELLDAYVMAAGVGS 779

QY 780 PYVSRLLIGLITSTVQLVLTQMPYGGCLLHVRENRGLCSQDLNMCQIANKMSYLEDV 839
 Db 780 PYVSRLLIGLITSTVQLVLTQMPYGGCLLHVRENRGLCSQDLNMCQIANKMSYLEDV 839
 QY 840 RLVRDLAARNVAVKSPNHYKIDFGIARLLDIDETEVADGKVPKIMMALESILRRF 899
 Db 840 RLVRDLAARNVAVKSPNHYKIDFGIARLLDIDETEVADGKVPKIMMALESILRRF 899
 QY 900 THQSDVMSYGVTVWEIMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMIMVXCM 959
 Db 900 THQSDVMSYGVTVWEIMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMIMVXCM 959
 QY 960 MIDSECRPRELIVSEFSRMAPRPFVVIQWEDIGPASPLDSTYRSLLEDDEMDGLVD 1019
 Db 960 MIDSECRPRELIVSEFSRMAPRPFVVIQWEDIGPASPLDSTYRSLLEDDEMDGLVD 1019
 QY 1020 AEEYLVPOQGFPCPDPAFGAGVYHRRHRSSTRSGGDLTLGLEPSEBEAPRSLAPSE 1079
 Db 1020 AEEYLVPOQGFPCPDPAFGAGVYHRRHRSSTRSGGDLTLGLEPSEBEAPRSLAPSE 1079
 QY 1080 GAGSDVFDGDLGMAKAGIQSLPTHDPSPLORYSEDPVPLPSETDGYVAPLTCSPQPEY 1139
 Db 1080 GAGSDVFDGDLGMAKAGIQSLPTHDPSPLORYSEDPVPLPSETDGYVAPLTCSPQPEY 1139
 QY 1140 VNOPDVPRQPPSPREGPLPAAPAPGATLERAKTLSPGKNGVYKQVFARFAGAVENPEYLT 1199
 Db 1140 VNOPDVPRQPPSPREGPLPAAPAPGATLERAKTLSPGKNGVYKQVFARFAGAVENPEYLT 1199
 QY 1200 QGGAPAPQHPPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGIDVPV 1255
 Db 1200 QGGAPAPQHPPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGIDVPV 1255

RESULT 2
 AAEL12130
 ID AAEL12130 standard; Protein; 1255 AA.
 AC AAEL12130;
 DT 18-DEC-2001 (first entry)
 DE Human tyrosine kinase-type receptor, HER-2.
 KW Therapeutic compound; major histocompatibility complex; vaccine;
 KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
 KW anti-cancer therapy; anti-cancer; breast cancer antigen; Apc;
 KW antigen presenting cell; human; tyrosine kinase-type receptor.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Region 774..782
 FT /note="Antigenic epitope"
 PN MO200168677-A2.
 PD 20-SEP-2001.
 PF 16-MAR-2001; 2001MCO-US40328.
 PR 16-MAR-2000; 2000US-0527487.
 PA (GENZ) GENZYME CORP.
 PI Nicotlette CA;
 DR WPI; 2001-616284/71.
 DR N-PSDB; AAD19731.
 XX Novel synthetic therapeutic compound for inducing immune response and
 PT for use in adoptive immunotherapy, has enhanced binding to major
 PT histocompatibility molecules and enhanced immunoregulatory properties

XX
 PS Claim 4; Page 63-67; 69pp; English.
 XX
 CC The invention relates to synthetic therapeutic compounds (antigenic
 CC peptides) with enhanced binding to major histocompatibility complex
 CC (MHC) molecules and enhanced immunoregulatory properties relative
 CC to their natural counterparts. Compounds of the invention are useful
 CC for inducing an immune response in a subject and for use in adoptive
 CC immunotherapy. They are useful as components of anti-cancer vaccines
 CC and to expand immune effector cells that are specific for cancers
 CC characterised by expression of the breast cancer antigen, HER-2.
 CC Polynucleotides that encode peptides of the invention are useful as
 CC hybridisation probes and as primers for the detection of genes of gene
 CC transcripts that are expressed in antigen presenting cells (APCs), to
 CC confirm transduction of polynucleotides into host cells. The present
 CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
 CC of the invention are designed based on the HER-2 antigenic peptide
 CC (774-782).
 SQ Sequence 1255 AA;
 XX
 Query Match 97.5%; Score 6646; DB 22; Length 1255;
 Best Local Similarity 97.7%; Pred. No. 0;
 Matches 1227; Conservative 7; Mismatches 20; Indels 2; Gaps 2;
 QY 1 MEIALACRMGLLALLPPGASTQVCTGDMKRLRPAPETHLDMRLRYGCGVQGNL 60
 Db 1 MEIALACRMGLLALLPPGASTQVCTGDMKRLRPAPETHLDMRLRYGCGVQGNL 60
 QY 61 ELTYLPTNASLSEFLDIOIEQGVYLIANQVROVPLQRLIVRGTLFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLSEFLDIOIEQGVYLIANQVROVPLQRLIVRGTLFEDNYALAVLDNG 120
 QY 121 DPLNNTTPYTGASPGGLRELOLRSLTELKXGVLIORNPOLCYQDTILMKOIFHKNOOLA 180
 Db 121 DPLNNTTPYTGASPGGLRELOLRSLTELKXGVLIORNPOLCYQDTILMKOIFHKNOOLA 180
 QY 122 DPLNNTTPYTGASPGGLRELOLRSLTELKXGVLIORNPOLCYQDTILMKOIFHKNOOLA 180
 Db 122 DPLNNTTPYTGASPGGLRELOLRSLTELKXGVLIORNPOLCYQDTILMKOIFHKNOOLA 180
 QY 181 LTLIDTNRBRACHPCSPWCKSGSRGWSESSDQSLTRTVACGACARCGPLPTDCHEQC 240
 Db 181 LTLIDTNRBRACHPCSPWCKSGSRGWSESSDQSLTRTVACGACARCGPLPTDCHEQC 240
 QY 241 AAGCTGPRNSDCLALHNHNSGICELHPALVYNTDFFSEMPREGRYTTGASCVTACP 300
 Db 241 AAGCTGPRNSDCLALHNHNSGICELHPALVYNTDFFSEMPREGRYTTGASCVTACP 300
 QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLMQYIKANSKFIGIT 360
 Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLMQYIKANSKFIGIT 360
 QY 361 ELFEAGCKKIFGSLAFIPESFDGPASTAFLOEBOQVFEETLEITGYLYISAMPDPLP 420
 Db 361 ELFEAGCKKIFGSLAFIPESFDGPASTAFLOEBOQVFEETLEITGYLYISAMPDPLP 420
 QY 421 DLSVFQNLQVIRGRILNHGAYSLTLOGIGISWGLRSLRELSGLAIHHNTHLCFHTV 480
 Db 421 DLSVFQNLQVIRGRILNHGAYSLTLOGIGISWGLRSLRELSGLAIHHNTHLCFHTV 480
 QY 481 PWDQLEFNPQALLHTANRPEDECVEGGLACHOLCAHCGMGPPTQCVNCSQFLRQDEC 540
 Db 481 PWDQLEFNPQALLHTANRPEDECVEGGLACHOLCAHCGMGPPTQCVNCSQFLRQDEC 540
 QY 541 VEECRVLOGLPREYVNAHCLPCHPECOQNGSYTCGPRADQCVACAHKDPFVCYARC 600
 Db 541 VEECRVLOGLPREYVNAHCLPCHPECOQNGSYTCGPRADQCVACAHKDPFVCYARC 600
 QY 601 PSQVAPDLISYPIWKFPDEBGAQPCPCINCTHSYVDLDDGCAEGRASLTISVAVVG 660
 Db 601 PSQVAPDLISYPIWKFPDEBGAQPCPCINCTHSYVDLDDGCAEGRASLTISVAVVG 660
 QY 661 ILTVVVLGVVFGILIKRQOKIRKYTRKRLLOETLEVEPLTPSGAMPNOQMLKEOYI 720
 Db 661 ILTVVVLGVVFGILIKRQOKIRKYTRKRLLOETLEVEPLTPSGAMPNOQMLKEOYI 720

QY 721 KANSKFTGTEL-TVYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDENYVAGVGS 779
 DB 721 R-KVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDENYVAGVGS 779
 QY 780 PYVSRLLGICLTSTVQVLTQMLPMYGCILDHVENRRLSGODLLMCMQIAKMSYLEDV 839
 DB 780 PYVSRLLGICLTSTVQVLTQMLPMYGCILDHVENRRLSGODLLMCMQIAKMSYLEDV 839
 QY 840 RLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGSKYPIKMALESILRRRF 899
 DB 840 RLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGSKYPIKMALESILRRRF 899
 QY 900 THOSDWSYGVATWELMTGKAPYDGIIPAREIPDLLEKGERLPOPPICIDIVYIMVCM 959
 DB 900 THOSDWSYGVATWELMTGKAPYDGIIPAREIPDLLEKGERLPOPPICIDIVYIMVCM 959
 QY 960 MIDSECRPRELIVSEFSRWARDPQRFVVIQNEIDLGPASPLDSTFYRSILLEDMDGLVD 1019
 DB 960 MIDSECRPRELIVSEFSRWARDPQRFVVIQNEIDLGPASPLDSTFYRSILLEDMDGLVD 1019
 QY 1020 AEEYLVPOQGFPCDDPAPGAGVWHHRHSSSTRSGGDLTLGLEPSEBEARSPPLAPSE 1079
 DB 1020 AEEYLVPOQGFPCDDPAPGAGVWHHRHSSSTRSGGDLTLGLEPSEBEARSPPLAPSE 1079
 QY 1080 GAGSDVFDGDLGMAKAGLSLPTHDPSPLOKXSEDPVPLPSETDGYVAPLTCSPQPEY 1139
 DB 1080 GAGSDVFDGDLGMAKAGLSLPTHDPSPLOKXSEDPVPLPSETDGYVAPLTCSPQPEY 1139
 QY 1140 VNQPDVFPQPPSPREGPLPAPAPAGATLERAKTLSFGKNGYVKDVAFEGAVENBEYLTP 1199
 DB 1140 VNQPDVFPQPPSPREGPLPAPAPAGATLERAKTLSFGKNGYVKDVAFEGAVENBEYLTP 1199
 QY 1200 QGGAAPQHPHPAPSPAFDNLVYWDQDPERGAPSTFGKTPAENBEYLGLDVEY 1255
 DB 1200 QGGAAPQHPHPAPSPAFDNLVYWDQDPERGAPSTFGKTPAENBEYLGLDVEY 1255

RESULT 3
 AAB60167 standard; Protein; 1255 AA.

XX AAB60167;
 AC AAB60167;
 DT 03-APR-2001 (first entry)
 DE HER2 transgene plasmid construct encoded protein.
 KM Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
 KM antibody.
 OS Homo sapiens.
 OS Synthetic.
 PN WO200100244-A2.
 PD 04-JAN-2001.
 PF 23-JUN-2000; 2000WO-US17229.
 XX 25-JUN-1999; 99US-0141316.
 PR 16-MAR-2000; 2000US-0189844.
 PA (GETH) GENENTECH INC.
 PI Erickson S, Schwall R;
 DR WPI; 2001-061962/07.
 DR N-PSDB; AAF24297.
 XX Treating tumors, particularly breast cancers, which overexpress an ErbB
 PT receptor and does not respond to an anti-ErbB antibody, comprises
 PT conjugating the antibody to a maytansinoid -
 XX

PS Example 3; Fig 4; 92pp; English.

XX The present invention provides a method of treating cancer by
 CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
 CC particular, the antibody is directed against ErbB2 (also known as HER2
 CC and p185neu). The method is particularly useful in the treatment of
 CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
 CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
 XX Sequence 1255 AA:
 QY
 Query Match 97.5%; Score 6646; DB 22; Length 1255;
 Best Local Similarity 97.7%; Pred. No. 0;
 Matches 1227; Conservative 7; Mismatches 20; Indels 2; Gaps 2

QY 1 MELAALCRNGLLALLPFGAASQVCTGDMKRLRPASETHLMDLRHLYOCQVYQNL 60
 DB 1 MELAALCRNGLLALLPFGAASQVCTGDMKRLRPASETHLMDLRHLYOCQVYQNL 60
 QY 61 ELTYLPTNASLSELDIQEVQGVLIANQVQVPLQRLIVRGQTLPEDNYALAVLNG 120
 DB 61 ELTYLPTNASLSELDIQEVQGVLIANQVQVPLQRLIVRGQTLPEDNYALAVLNG 120
 QY 121 DPLNNTPTVTGASPGGLRLQLRSLTEILKGVLLQRNPQLCYQPTILMKDIFHKMOLA 180
 DB 121 DPLNNTPTVTGASPGGLRLQLRSLTEILKGVLLQRNPQLCYQPTILMKDIFHKMOLA 180
 QY 181 LTLIDNRSRACPCSPMKGSRGWSSESDCSLTRTYCAGGACRCKPLPTDCHEQC 240
 DB 181 LTLIDNRSRACPCSPMKGSRGWSSESDCSLTRTYCAGGACRCKPLPTDCHEQC 240
 QY 241 AAGCTGKXSDCLACJLHFNHSGICELHCPALVYNTDTFESMPNBRGYTFASCTYAC 300
 DB 241 AAGCTGKXSDCLACJLHFNHSGICELHCPALVYNTDTFESMPNBRGYTFASCTYAC 300
 QY 301 YNYLSTVSGCTLVCPHNOEYVAEFGTCRCKSPCARVYCGIMQVIXANSKPIGTT 360
 DB 301 YNYLSTVSGCTLVCPHNOEYVAEFGTCRCKSPCARVYCGIMQVIXANSKPIGTT 360
 QY 361 ELERAGCKIFGSLAFIPESFDDPASNTAPLOPELOVFEETLGYLISAWPDSL 420
 DB 361 ELERAGCKIFGSLAFIPESFDDPASNTAPLOPELOVFEETLGYLISAWPDSL 420
 QY 421 DLSYFQNLQVIRGRILHNAGVSLTLQGLISWGLSLSELGSGALLHHNHLCTVHTV 480
 DB 421 DLSYFQNLQVIRGRILHNAGVSLTLQGLISWGLSLSELGSGALLHHNHLCTVHTV 480
 QY 481 PMDOLFENPHQALLHTANRPEDECEVGEGLACHQLCARGHCMWGPPTQCVNCSQFLRGQEC 540
 DB 481 PMDOLFENPHQALLHTANRPEDECEVGEGLACHQLCARGHCMWGPPTQCVNCSQFLRGQEC 540
 QY 541 VEECRVLIQSLPREYVNAARHCLPHPECOQONGSVTCFGEADOCVACAHYKDPFVCVANC 600
 DB 541 VEECRVLIQSLPREYVNAARHCLPHPECOQONGSVTCFGEADOCVACAHYKDPFVCVANC 600
 QY 601 PSGVKPLSLMPTMKPDEDEGACOPCPINCTHSCVDLDKGCAPABORASPLTSISAVVG 660
 DB 601 PSGVKPLSLMPTMKPDEDEGACOPCPINCTHSCVDLDKGCAPABORASPLTSISAVVG 660
 QY 661 ILLVVLGVVFGILLIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNOAQRILKEQYI 720
 DB 661 ILLVVLGVVFGILLIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNOAQRILKEQYI 720
 QY 721 KANSKFTGTEL-TVYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDENYVAGVGS 779
 DB 721 KANSKFTGTEL-TVYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDENYVAGVGS 779
 QY 780 PYVSRLLGICLTSTVQVLTQMLPMYGCILDHVENRRLSGODLLMCMQIAKMSYLEDV 839
 DB 780 PYVSRLLGICLTSTVQVLTQMLPMYGCILDHVENRRLSGODLLMCMQIAKMSYLEDV 839
 QY 840 RLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGSKYPIKMALESILRRRF 899
 DB 840 RLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGSKYPIKMALESILRRRF 899

Db 840 RLVRHDLAARNVLVKSPPNKHITDFGLARLLIDIDETEVHADGCKVPIKMALESILRRR 899
QY THSDPWSYGVTVNELMTFGAKPYDGI PAREIPDLLEKGERLPPQPICTIDVYMIWKCM 959
Db THSDPWSYGVTVNELMTFGAKPYDGI PAREIPDLLEKGERLPPQPICTIDVYMIWKCM 959
QY 960 MIDSECRPFRELVSSEFRMARDPQRFVIVIONEDLGPASPLDSTFFRSLLIEDDDMDLVD 1019
Db 960 MIDSECRPFRELVSSEFRMARDPQRFVIVIONEDLGPASPLDSTFFRSLLIEDDDMDLVD 1019
QY 1020 AEEYLVPPQGFCCDPAPGAGAMVHHRRSSSTBSGGGDLTLIGLEPSEEBAPSPPLAPSE 1079
Db 1020 AEEYLVPPQGFCCDPAPGAGAMVHHRRSSSTBSGGGDLTLIGLEPSEEBAPSPPLAPSE 1079
QY 1080 GAGSDVFPDGLGMAKGLQSLPTHDPSPLOQYSEDPVLPJSETDGYVAPLITCSPQPEX 1139
Db 1080 GAGSDVFPDGLGMAKGLQSLPTHDPSPLOQYSEDPVLPJSETDGYVAPLITCSPQPEX 1139
QY 1140 VNQPDVRRPQSPSPREGPLPAPRAGATLERAKTILSPGKNGVWDVAFGAVENPEYLTP 1199
Db 1140 VNQPDVRRPQSPSPREGPLPAPRAGATLERAKTILSPGKNGVWDVAFGAVENPEYLTP 1199
QY 1200 QGGAAPQHPPEPASPAPNDLYWDDPPEKGAAPPSTFKGTPTAENPEYLGIDVPV 1255
Db 1200 QGGAAPQHPPEPASPAPNDLYWDDPPEKGAAPPSTFKGTPTAENPEYLGIDVPV 1255

RESULT 4
AAU74545
ID AAU74545 standard; Protein; 1255 AA.
AC AAU74545;
DT 23-APR-2002 (first entry)
XX Human HER2 (ErbB2) polypeptide.
DE Human HER2 (ErbB2) polypeptide.
XX
XX Human; HER2; ErbB; epidermal growth factor receptor; receptor;
KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
KW glioma disorder; astrocytoma disorder; hypothalamic disorder;
KW glandular disorder; macrophage disorder; epithelial disorder;
KW stromal disorder; blastocoele disorder; inflammatory disorder;
KW angiogenic disorder; immunological disorder.
XX
XX Homo sapiens.
OS
XX US2002001587-A1.
PN 03-JAN-2002.
PD 16-MAR-2001; 2001US-0811123.
PE 16-MAR-2000; 2000US-189844P.
PR 05-OCT-2000; 2000US-238327P.
XX
XX (ERIC/) ERICKSON S.
PA (SCHW/) SCHWALL R.
PA (SLIW/) SLIWOMSKI M.
XX
XX Erickson S, Schwall R, Sliwowski M;
PI WPI; 2002-163686/21.
DR N-PSDB; ABK14058.
XX
XX Treating tumour characterised by overexpression of epidermal growth
PT factor receptor. ErbB or cancer in mammal, comprises administering
PT anti-ErbB antibody-maytansinoid conjugate to the mammal -
XX
XX Example 3; Fig 7; 33pp; English.
XX
XX The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
CC receptor (ErbB) and does not respond or responds poorly, to treatment
CC with an anti-ErbB antibody, comprising administering to the mammal an
CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
CC prostate and bladder, preferably breast cancer. The breast cancer is a
CC metastatic breast cancer or an aggressive form of metastatic breast
CC cancer which overexpresses ErbB2. The method is also useful for treating
CC neuronal, glial, astrocytic, hypothalamic, glandular, macrophage,
CC epithelial, stromal, blastocoele, inflammatory, angiogenic and
CC immunological disorders. This sequence represents the human HER2 (ErbB2)
CC polypeptide of the invention.
XX
SQ Sequence 1255 AA;
Query Match 97.5%; Score 6646; DB 23; Length 1255;
Best Local Similarity 97.7%; Pred. No. 0;
Matches 1227; Conservative 7; Mismatches 20; Indels 2; Gaps 2;
QY 1 MELALCRWGLLALLPPGAASITQCTGDMKRLPASPETHLDMRLHYQSGQVQGNL 60
Db 1 MELALCRWGLLALLPPGAASITQCTGDMKRLPASPETHLDMRLHYQSGQVQGNL 60
QY 61 ELTYLPTNASTLFIQDIQEVQGYVLIHANOVRQVPLQRLIRVGTQLFEDNVYALAVLDNG 120
Db 61 ELTYLPTNASTLFIQDIQEVQGYVLIHANOVRQVPLQRLIRVGTQLFEDNVYALAVLDNG 120
QY 121 DPLNNTTPVTCASGSGRLRELQRLSLTEILKGVLIQBNPOLCYQDTILMKDITHKNNQLA 180
Db 121 DPLNNTTPVTCASGSGRLRELQRLSLTEILKGVLIQBNPOLCYQDTILMKDITHKNNQLA 180
QY 181 LTLIDITNRSRACHDPCSPCKSGKSCWSSSEDCQSLRTVTCAGGACARCKGPLPTDCHEQC 240
Db 181 LTLIDITNRSRACHDPCSPCKSGKSCWSSSEDCQSLRTVTCAGGACARCKGPLPTDCHEQC 240
QY 241 AAGCTGPRHSDCLALCFPNHSGICELHCPALVYVNDTFESMNPBGRYVFGASCYTACP 300
Db 241 AAGCTGPRHSDCLALCFPNHSGICELHCPALVYVNDTFESMNPBGRYVFGASCYTACP 300
QY 301 YNYSTVGSCTIACPLHNOBVTAEADGTQRCCKSKPCARVCYGLQWQYKANSKFTIGIT 360
Db 301 YNYSTVGSCTIACPLHNOBVTAEADGTQRCCKSKPCARVCYGLQWQYKANSKFTIGIT 360
QY 361 ELBRAGCKTIFGSLAFIPESFDGDPASNTAPLOEQVQVETLEITGYLYISAMPDSLIP 420
Db 361 IQEPAGCKTIFGSLAFIPESFDGDPASNTAPLOEQVQVETLEITGYLYISAMPDSLIP 420
QY 421 DLSVFONLQVTRGRILHNGAVSLTLQGLGSMGLSLRELSSGLALIHNTHLCEVHTV 480
Db 421 DLSVFONLQVTRGRILHNGAVSLTLQGLGSMGLSLRELSSGLALIHNTHLCEVHTV 480
QY 481 PMDQLFRNPQALHTANRPEDCEVSGELACHOLCARGHCWGPPTQCVNCSOFLRGQEC 540
Db 481 PMDQLFRNPQALHTANRPEDCEVSGELACHOLCARGHCWGPPTQCVNCSOFLRGQEC 540
QY 541 VEECRVLOGLPREVYVNRHCLPHPEQOPONGSVTCFGBPADCCVACAHXKDPFCVANC 600
Db 541 VEECRVLOGLPREVYVNRHCLPHPEQOPONGSVTCFGBPADCCVACAHXKDPFCVANC 600
QY 601 PSQVPLSLVYPIKFPDEGACOPCEINTHSCVDLDKGCAPBQASPLTIVSAVVG 660
Db 601 PSQVPLSLVYPIKFPDEGACOPCEINTHSCVDLDKGCAPBQASPLTIVSAVVG 660
QY 661 ILTVVVLGVVFGILIKRQOKIRKRYTRRLQETELVEPLTPSGAMPNOQWILKTEL 720
Db 661 ILTVVVLGVVFGILIKRQOKIRKRYTRRLQETELVEPLTPSGAMPNOQWILKTEL 720
QY 721 KANSKFTIGITEL-TVYGIWIPGENVKIPVAKVLRNTSPKANKILDEAVYVAGVS 779
Db 721 R-KVTVIGSGAFGVTVGIWIPGENVKIPVAKVLRNTSPKANKILDEAVYVAGVS 779
QY 780 PYVSRLLIGICLTSTVQLVLTQJLMPYGLCLDHVRENRLGSLQDILNMCQIAGKMSYLEDP 839

```

Db      780 PYVSRLLGICLTSTVQLVTQLMPYGCGLDHRVNRGRGLSGQDILLNMCQIAKCMSTLEEV 839
Qy      840 FLVHRDLAARNVLYKSPNHVKITDFGLARLLIDETETAHGDKRPVKMALESIIRRF 899
Db      840 FLVHRDLAARNVLYKSPNHVKITDFGLARLLIDETETAHGDKRPVKMALESIIRRF 899
Qy      900 THOSDWSYGVTVLMTFGAKPYDGIIPABEIPDLLEKGRLLPQPICTIDVYMIWKCM 959
Db      900 THOSDWSYGVTVLMTFGAKPYDGIIPABEIPDLLEKGRLLPQPICTIDVYMIWKCM 959
Qy      960 MIDSECRFRFRELVSFSRMAADPQRFVITQNEDLGPASPLDSTFYRSLLEDMDMDLVD 1019
Db      960 MIDSECRFRFRELVSFSRMAADPQRFVITQNEDLGPASPLDSTFYRSLLEDMDMDLVD 1019
Qy      1020 AEEYLVPOQGFCECPDPAAGGWMVHHRRSSSTRSGGDLTLGLESEEBAPRSLAPSE 1079
Db      1020 AEEYLVPOQGFCECPDPAAGGWMVHHRRSSSTRSGGDLTLGLESEEBAPRSLAPSE 1079
Qy      1080 GAGSDVPDGLGMAAGLQSLPTHDSPLQRYSEDPVLPSETDGYVAPLTCSPOPEY 1139
Db      1080 GAGSDVPDGLGMAAGLQSLPTHDSPLQRYSEDPVLPSETDGYVAPLTCSPOPEY 1139
Qy      1140 VNQPDVPRQPPSPREGPLPAARPAAGATLERAKTSLSPGKGVADVPAFGAVENPEYLTP 1199
Db      1140 VNQPDVPRQPPSPREGPLPAARPAAGATLERAKTSLSPGKGVADVPAFGAVENPEYLTP 1199
Qy      1200 QGGAAPQHPPPAPSPAFDNLVYWDQPPERGAAPPSTFKGPTAENPEYLGDPV 1255
Db      1200 QGGAAPQHPPPAPSPAFDNLVYWDQPPERGAAPPSTFKGPTAENPEYLGDPV 1255

RESULT 5
AA01111 ID AA01111 standard; Protein; 1255 AA.
AC      AA01111;
DT      01-JAN-1997 (first entry)
DE      HER-2/neu protein.
XX      HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW      breast cancer; ovary cancer; colon cancer; lung cancer;
KM      prostate cancer; immunisation; tumour; vaccine; vector.
OS      Homo sapiens.
XX      Key Location/Qualifiers
FT      676..1255
FT      /label= "intracellular domain
FT      /note= "claimed domain, useful for immunisation"
XX      MO9630514-A1.
XX      03-OCT-1996.
XX      28-MAR-1996; 96MO-US01689.
XX      31-MAR-1995; 95US-0414417.
XX      (UNIW ) UNIV WASHINGTON.
XX      Cheever MA, Disis ML;
XX      MPI: 1996-455361/45.
XX      N-PSDB; AAT40739.
XX      DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
XX      treatment of malignancies with which the HER-2/neu oncogene is
XX      associated
XX      Claim 2; 'page 56-61; 71pp; English.

```

```

XX      Human HER-2/neu protein (AA01111), also called p185 or c-erbB2, is
CC      the product of the HER-2/neu oncogene (see also AAT40739).. The
CC      protein is over-expressed in various cancers, including breast,
CC      ovarian, colon, lung and prostate. The intracellular domain of the
CC      protein can be used to immunise an animal against a malignancy with
CC      which the oncogene is associated. The polypeptide can be produced
CC      in transfected host cells for use in immunisation. Alternatively,
CC      animal cells are transfected in vivo or ex vivo with a viral vector
CC      that directs expression of the polypeptide.
XX      Sequence 1255 AA:
SO      Query Match 97.4%; Score 6640; DB 17; Length 1255;
        Best Local Similarity 97.5%; Pred. No. 0;
        Matches 1225; Conservative 0; Mismatches 21; Indels 2; Gaps 2;
Qy      1 MELAALCRWGLLALLPFGAASVQVCTGTDMLRLPASETHLDMRLHYQGCQVVGNTL 60
Db      1 MELAALCRWGLLALLPFGAASVQVCTGTDMLRLPASETHLDMRLHYQGCQVVGNTL 60
Qy      61 ELTYLPNASTLFDITQEVQGVLLIAHNOVRQVPLQRIRVGTQLPEDNYALAVLDNG 120
Db      61 ELTYLPNASTLFDITQEVQGVLLIAHNOVRQVPLQRIRVGTQLPEDNYALAVLDNG 120
Qy      121 DPLNNTTPTVGASPGGLRELQRLSTEILKGVLLIQRNQLCYQDTILMKDIFHKNNQLA 180
Db      121 DPLNNTTPTVGASPGGLRELQRLSTEILKGVLLIQRNQLCYQDTILMKDIFHKNNQLA 180
Qy      181 LTLIDNTRSPACHPCSPCKSGSKGSESSDCQSLTRYCAGGACARCPRLPTDCHDEC 240
Db      181 LTLIDNTRSPACHPCSPCKSGSKGSESSDCQSLTRYCAGGACARCPRLPTDCHDEC 240
Qy      181 LTLIDNTRSPACHPCSPCKSGSKGSESSDCQSLTRYCAGGACARCPRLPTDCHDEC 240
Db      181 LTLIDNTRSPACHPCSPCKSGSKGSESSDCQSLTRYCAGGACARCPRLPTDCHDEC 240
Qy      241 AAGCTGPKHSDCLACHFNHSGICEIHCPALVTYNTDTPESMPNPEGRTPFGSCVTACP 300
Db      241 AAGCTGPKHSDCLACHFNHSGICEIHCPALVTYNTDTPESMPNPEGRTPFGSCVTACP 300
Qy      241 AAGCTGPKHSDCLACHFNHSGICEIHCPALVTYNTDTPESMPNPEGRTPFGSCVTACP 300
Db      241 AAGCTGPKHSDCLACHFNHSGICEIHCPALVTYNTDTPESMPNPEGRTPFGSCVTACP 300
Qy      301 YNYLSTDVSGCTIVCPILHNGEVTAEADGTORCEKSKPCARVCYGLQMOYIKANSKFTGIT 360
Db      301 YNYLSTDVSGCTIVCPILHNGEVTAEADGTORCEKSKPCARVCYGLQMOYIKANSKFTGIT 360
Qy      361 ELFPAGCKKIFGSLAFPSFDDPSNTPAPQBPQVFTLEETIGLYISAMPDLSIP 420
Db      361 ELFPAGCKKIFGSLAFPSFDDPSNTPAPQBPQVFTLEETIGLYISAMPDLSIP 420
Qy      361 IQEPAGCKKIFGSLAFPSFDDPSNTPAPQBPQVFTLEETIGLYISAMPDLSIP 420
Db      361 IQEPAGCKKIFGSLAFPSFDDPSNTPAPQBPQVFTLEETIGLYISAMPDLSIP 420
Qy      421 DLSYFQNLQVIRGRILHNGAYSLTLQGLISWLGRLSRELSGIALIHNHHLGCVHTV 480
Db      421 DLSYFQNLQVIRGRILHNGAYSLTLQGLISWLGRLSRELSGIALIHNHHLGCVHTV 480
Qy      481 PMDOLFNPQHALLHTANRPEDCEVGEGLACHQLCARGHCMGPGPTQCVNGSQFLRGQEC 540
Db      481 PMDOLFNPQHALLHTANRPEDCEVGEGLACHQLCARGHCMGPGPTQCVNGSQFLRGQEC 540
Qy      481 PMDOLFNPQHALLHTANRPEDCEVGEGLACHQLCARGHCMGPGPTQCVNGSQFLRGQEC 540
Db      481 PMDOLFNPQHALLHTANRPEDCEVGEGLACHQLCARGHCMGPGPTQCVNGSQFLRGQEC 540
Qy      541 VEEGRVLYQGLPREVYVARHCLPCHPCQFQNGSVTCFGEADQCAAHYKDPFCVARG 600
Db      541 VEEGRVLYQGLPREVYVARHCLPCHPCQFQNGSVTCFGEADQCAAHYKDPFCVARG 600
Qy      541 VEEGRVLYQGLPREVYVARHCLPCHPCQFQNGSVTCFGEADQCAAHYKDPFCVARG 600
Db      541 VEEGRVLYQGLPREVYVARHCLPCHPCQFQNGSVTCFGEADQCAAHYKDPFCVARG 600
Qy      601 PSQVYKPLSTYMPIMKPPDEGACQPCPINCTHSQVLDLQKGPAPQASPLTSISAVVG 660
Db      601 PSQVYKPLSTYMPIMKPPDEGACQPCPINCTHSQVLDLQKGPAPQASPLTSISAVVG 660
Qy      601 PSQVYKPLSTYMPIMKPPDEGACQPCPINCTHSQVLDLQKGPAPQASPLTSISAVVG 660
Db      601 PSQVYKPLSTYMPIMKPPDEGACQPCPINCTHSQVLDLQKGPAPQASPLTSISAVVG 660
Qy      661 ILVYVLYGVVFGILLIRROOKIRKYMRRLLQSTELVEPLTPSGAMPNQAQRILKETEL 720
Db      661 ILVYVLYGVVFGILLIRROOKIRKYMRRLLQSTELVEPLTPSGAMPNQAQRILKETEL 720
Qy      721 KANSKFTIGTEL-TYVKGIWIPDGENYKIPVAIKVLRNTSPKANKELIDEAYVAGVGS 779
Db      721 KANSKFTIGTEL-TYVKGIWIPDGENYKIPVAIKVLRNTSPKANKELIDEAYVAGVGS 779
Qy      721 KANSKFTIGTEL-TYVKGIWIPDGENYKIPVAIKVLRNTSPKANKELIDEAYVAGVGS 779
Db      721 KANSKFTIGTEL-TYVKGIWIPDGENYKIPVAIKVLRNTSPKANKELIDEAYVAGVGS 779
Qy      780 PYVSRLLGICLTSTVQLVTQLMPYGCGLDHRVNRGRGLSGQDILLNMCQIAKCMSTLEEV 839
Db      780 PYVSRLLGICLTSTVQLVTQLMPYGCGLDHRVNRGRGLSGQDILLNMCQIAKCMSTLEEV 839

```



```

QY 840 RLVHRDLAARNVLYKSPNHVKITDPEGLARLDIDETEVHADGKVPITKMALESILRRRF 899
DB 840 RLVHRDLAARNVLYKSPNHVKITDPEGLARLDIDETEVHADGKVPITKMALESILRRRF 899
QY 900 THQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMWKW 959
DB 900 THQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMWKW 959
QY 960 MIDSECRPRFRELVSFSPRARDPORFVVIQNEIDLGPASPLDSTYRSLDDDMGLVD 1019
DB 960 MIDSECRPRFRELVSFSPRARDPORFVVIQNEIDLGPASPLDSTYRSLDDDMGLVD 1019
QY 1020 ABEYLVPQGFPCPDPAFGAGVWHRHRSSTRSQGGDLTGLPSESEAPRSLAPSE 1079
DB 1020 ABEYLVPQGFPCPDPAFGAGVWHRHRSSTRSQGGDLTGLPSESEAPRSLAPSE 1079
QY 1080 GAGSDVFDGLMGAAKGLQSLPTHPDPLQRYSEDPTVPLPSETDGYVALTCSPOPEY 1139
DB 1080 GAGSDVFDGLMGAAKGLQSLPTHPDPLQRYSEDPTVPLPSETDGYVALTCSPOPEY 1139
QY 1140 VNCPDVRPQPPSPREGPLPAPAPAGATLEAKTILSPGKGVKQVFAFGAVENPEYLTP 1199
DB 1140 VNCPDVRPQPPSPREGPLPAPAPAGATLEAKTILSPGKGVKQVFAFGAVENPEYLTP 1199
QY 1200 QGGAAPQPPHPPAPSPAFDNLVYWDODPPERGAPOSTFKGPTAENPEYLGIDVPV 1255
DB 1200 QGGAAPQPPHPPAPSPAFDNLVYWDODPPERGAPOSTFKGPTAENPEYLGIDVPV 1255

```

RESULT 6

AAW92406 standard; Protein: 1255 AA.

AAW92406;

21-APR-1999 (first entry)

Human HER-2/neu oncogene protein.

HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;

malignancy; treatment; tumour.

Homo sapiens.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

```

Key Location/Qualifiers
Region 676..1255
/ note="region which elicits immune response"

US5869445-A.
09-FEB-1999.
01-APR-1996; 96US-0625101.
01-APR-1996; 96US-0625101.
17-MAR-1993; 93US-0033644.
12-AUG-1993; 93US-0106112.
31-MAR-1995; 95US-0414417.
(UNIM ) UNIV WASHINGTON.
PI Cheever MA, Disis ML;
XX WPI, 1999-152835/13.
XX N-PSDB; AAX01912.
XX Use of HER-2/neu polypeptides - for eliciting an immune response to
XX an HER-2/neu associated malignancy, particularly for treating or
XX preventing tumours
XX Claim 3; Column 31-38; 26pp; English.
XX This sequence represents the human HER-2/neu oncogene protein. A fragment
CC

```

of this protein is used in a method for eliciting or enhancing an immune response to HER-2/neu protein. The polypeptide can stimulate T cells and B cells to produce an immune response to the HER-2/neu protein. The method can be used for immunisation against a malignancy in which the HER-2/neu oncogene is associated and in the treatment of an existing tumour, or to prevent tumour occurrence or reoccurrence.

Sequence 1255 AA;

Query Match 97.4%; Score 6640; DB 20; Length 1255;

Best Local Similarity 97.5%; Pred. No. 0;

Matches 1225; Conservative 8; Mismatches 21; Indels 2; Gaps 2

```

QY 1 MEALALCWGLLLALLPPGAASVTCGTGTMKRLPASPETHLDMRLHYOGQVQVQNL 60
DB 1 MEALALCWGLLLALLPPGAASVTCGTGTMKRLPASPETHLDMRLHYOGQVQVQNL 60
QY 61 ELTYLPNASISFLQDIQEVQGYLLAHNOVRQVPLQRLRYRGTLQFEDNALAVLDNG 120
DB 61 ELTYLPNASISFLQDIQEVQGYLLAHNOVRQVPLQRLRYRGTLQFEDNALAVLDNG 120
QY 121 DPLNNTTPTVQASPGSLRELQSLRELTLEKGYVLQRNPQLQVDTILMKDIFHKXNOLA 180
DB 121 DPLNNTTPTVQASPGSLRELQSLRELTLEKGYVLQRNPQLQVDTILMKDIFHKXNOLA 180
QY 181 LTLIDTNRSRACHPCSPMCKSGRCWGSSEDCQSLRTVCAAGGACCKGPLPTDCHEQC 240
DB 181 LTLIDTNRSRACHPCSPMCKSGRCWGSSEDCQSLRTVCAAGGACCKGPLPTDCHEQC 240
QY 241 AAGCTGPKHSDCLALCPHNSGICELCPALVYNTDTPESMNPREGRTFFGASCTACP 300
DB 241 AAGCTGPKHSDCLALCPHNSGICELCPALVYNTDTPESMNPREGRTFFGASCTACP 300
QY 301 YNYLSTDVGSCTLVCPHNOEVTAEQDTCCEKSPKACVCGTGLMEHLREYAVTSAN 360
DB 301 YNYLSTDVGSCTLVCPHNOEVTAEQDTCCEKSPKACVCGTGLMEHLREYAVTSAN 360
QY 361 ELFPAGCKITFGLAPLPESFDGPAANTPLQPELOQVETLEETGLYLSAMPDSLP 420
DB 361 ELFPAGCKITFGLAPLPESFDGPAANTPLQPELOQVETLEETGLYLSAMPDSLP 420
QY 421 DLSVFNQVIRGIRILHNGAYSLLTQGLISWGLSLRELGSGLALIHNNHTLCVHTV 480
DB 421 DLSVFNQVIRGIRILHNGAYSLLTQGLISWGLSLRELGSGLALIHNNHTLCVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDCEVGEGLACHQLCARHCWGGPQVCVCSQFLRGDEC 540
DB 481 PWDQLFRNPHQALLHTANRPEDCEVGEGLACHQLCARHCWGGPQVCVCSQFLRGDEC 540
QY 541 VEECRVLOGLPREVYNARHCLPCHPEQOPNGSVTCFEPADQCVACAHYKDPFCVABC 600
DB 541 VEECRVLOGLPREVYNARHCLPCHPEQOPNGSVTCFEPADQCVACAHYKDPFCVABC 600
QY 601 PSGVYKPDLSYMPFKFPDEBGAQCPICINCTHSCVDLDKCPAERASPLSTYSAYVG 660
DB 601 PSGVYKPDLSYMPFKFPDEBGAQCPICINCTHSCVDLDKCPAERASPLSTYSAYVG 660
QY 661 ILVYVGLGVFGIILIKRQOKIRKYTRKRLQSTELVEPLTPSGAVPNOAKRIKECYI 720
DB 661 ILVYVGLGVFGIILIKRQOKIRKYTRKRLQSTELVEPLTPSGAVPNOAKRIKECYI 720
QY 721 KANSKFIGITEL-TVYKGIWIPDGENYKIFVALIKVRENTSPANKELIDEAYVMAVGS 779
DB 721 KANSKFIGITEL-TVYKGIWIPDGENYKIFVALIKVRENTSPANKELIDEAYVMAVGS 779
QY 780 PYVSRLIGICLTSYVQLVTLQMPGCLDHYRENRRLSODLLANCMQIAGMSYLEVD 839
DB 780 PYVSRLIGICLTSYVQLVTLQMPGCLDHYRENRRLSODLLANCMQIAGMSYLEVD 839
QY 840 RLVHRDLAARNVLYKSPNHVKITDPEGLARLDIDETEVHADGKVPITKMALESILRRRF 899
DB 840 RLVHRDLAARNVLYKSPNHVKITDPEGLARLDIDETEVHADGKVPITKMALESILRRRF 899

```

QY 900 THSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLLEKGERLPPICTIDVYIMVWKCM 959
 Db 900 THSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLLEKGERLPPICTIDVYIMVWKCM 959
 QY 960 MIDSECRPRFRELVSSESRMARDPQRFVVIQNEIDLGPASFLDSTFYRSLLEDDMDGLVD 1019
 Db 960 MIDSECRPRFRELVSSESRMARDPQRFVVIQNEIDLGPASFLDSTFYRSLLEDDMDGLVD 1019
 QY 1020 AEEYLVPOQGFCCPDPAFGAGVHHRHSSSTRSGGDLTLGLEPSEEAERSLAPSE 1079
 Db 1020 AEEYLVPOQGFCCPDPAFGAGVHHRHSSSTRSGGDLTLGLEPSEEAERSLAPSE 1079
 QY 1080 GAGSDVFDGDLGMAAKGLQSLPTHDPSPLOQYSEDFTVPLPSETDGYVAPLTCSPQPEY 1139
 Db 1080 GAGSDVFDGDLGMAAKGLQSLPTHDPSPLOQYSEDFTVPLPSETDGYVAPLTCSPQPEY 1139
 QY 1140 VNQPDVRFQPPSPREGLPLAARPAAGATLERAKTSLSPGKGVGVKVFAPAGAVENPEYLTP 1139
 Db 1140 VNQPDVRFQPPSPREGLPLAARPAAGATLERAKTSLSPGKGVGVKVFAPAGAVENPEYLTP 1139
 QY 1200 QGGAPQPPHPPAPFSPAFDNLVYWDQDPPERGAPPSFTKGTPTAENPEYLGLDVPV 1255
 Db 1200 QGGAPQPPHPPAPFSPAFDNLVYWDQDPPERGAPPSFTKGTPTAENPEYLGLDVPV 1255
 RESULT 7
 AAB21198
 ID AAB21198 standard; protein; 1255 AA.
 AC AAB21198;
 XX
 XX 12-JAN-2001 (first entry)
 DT
 XX
 DE Human HER-2/neu protein.
 XX
 XX Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
 KM breast cancer; prostate cancer; ovarian cancer; lung cancer;
 KM colon cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200044895-A1.
 PD 03-AUG-2000.
 XX
 PF 28-JAN-2000; 2000WO-US02164.
 XX
 PR 29-JAN-1999; 99US-0117976.
 XX
 PA (CORI-) CORIXA CORP.
 XX (SMIK) SMITHKLINE BEECHAM.
 XX
 PI Cheever MA, Gheysen D;
 XX
 DR WPI: 2000-505976/45.
 DR N-PSDB; AAA89736.
 XX
 PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
 PT useful for vaccinating against breast, ovarian, colon, lung and
 PT prostate cancers -
 XX
 PS Claim 52; Fig 7; 128pp; English.
 XX
 CC The present sequence is the human HER-2/neu protein. It is a member of
 CC the tyrosine kinase family of receptor-like glycoproteins and shows
 CC homology to the epidermal growth factor receptor (EGFR). It probably
 CC plays a part in cell growth and/or differentiation. The HER-2/neu
 CC gene is an oncogene. An HER-2/neu fusion protein comprising a
 CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
 CC domain may be used to treat or prevent cancer by eliciting or
 CC enhancing an immune response to the HER-2/neu protein. It may be used
 CC to treat malignancies such as breast, ovarian, colon, lung and
 CC prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.
 XX
 SQ Sequence 1255 AA;
 Query Match 97.4%; Score 6640; DB 21; Length 1255;
 Best Local Similarity 97.5%; Pred No. 0;
 Matches 1225; Conservative 8; Mismatches 21; Indels 2; Gaps 2.
 QY 1 MELALCRMGLLALLPQGAASVQCTGDMKRLPASPETHLDMRLYQGCQVQGNL 60
 Db 1 MELALCRMGLLALLPQGAASVQCTGDMKRLPASPETHLDMRLYQGCQVQGNL 60
 QY 61 ELTYLPYTNASLFLDIOIEVQGVYLIANOVROYVLOSLRIVRGQVPEENYALAVLDNG 120
 Db 61 ELTYLPYTNASLFLDIOIEVQGVYLIANOVROYVLOSLRIVRGQVPEENYALAVLDNG 120
 QY 121 DPLNNTPTVYGASPGGLRELOLRSLTEILKGVYLIQRPOLCYODTILMKDIFKNNOLA 180
 Db 121 DPLNNTPTVYGASPGGLRELOLRSLTEILKGVYLIQRPOLCYODTILMKDIFKNNOLA 180
 QY 181 LTLIDTNSRACHPCSPMCKSGSRCKWSESSDCCSLTRTVCAAGGACARCKGLPTDCCHEQC 240
 Db 181 LTLIDTNSRACHPCSPMCKSGSRCKWSESSDCCSLTRTVCAAGGACARCKGLPTDCCHEQC 240
 QY 241 AAGCTGPKHSDCLACLHFNHSGICEFHCPALVYNTDTFESMPNREGRYTFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACLHFNHSGICEFHCPALVYNTDTFESMPNREGRYTFGASCVTACP 300
 QY 301 YNYLSTDVSGCTLVCPPLHNOGVTAEDGTORCEKSKPCARVYGLGNOYI KANSKFTGIT 360
 Db 301 YNYLSTDVSGCTLVCPPLHNOGVTAEDGTORCEKSKPCARVYGLGMEHREVAVTSAN 360
 QY 361 ELEFAGCKKIFGSLAFIPESFDGPGASTAPLOEOUOVETLEITGYLYISMPISLP 420
 Db 361 IOEFAGCKKIFGSLAFIPESFDGPGASTAPLOEOUOVETLEITGYLYISMPISLP 420
 QY 421 DLSVFQNLQVIRGRILNHGAYSLTLQGLISWLSLRLSRLREISGLALIHNTHLCEFTV 480
 Db 421 DLSVFQNLQVIRGRILNHGAYSLTLQGLISWLSLRLSRLREISGLALIHNTHLCEFTV 480
 QY 481 PWDQFRPHQALLHTARPEDECVGGLACHQICAGHCGMGPGPTCCVNCOSFLRQEC 540
 Db 481 PWDQFRPHQALLHTARPEDECVGGLACHQICAGHCGMGPGPTCCVNCOSFLRQEC 540
 QY 541 VEECRVLOGLPREVYNARHCLPCHPECOPQNGSVTCGFPADCCVACAHKDDPFVYARC 600
 Db 541 VEECRVLOGLPREVYNARHCLPCHPECOPQNGSVTCGFPADCCVACAHKDDPFVYARC 600
 QY 601 PSQVYKPDLSYMPIMKFPDEBGAQPCPCINCTHSCVDLDDKGCFAEORASPLTISAVVG 660
 Db 601 PSQVYKPDLSYMPIMKFPDEBGAQPCPCINCTHSCVDLDDKGCFAEORASPLTISAVVG 660
 QY 661 ILLVVLGVVFGILIKSRQOKIRYTRRSLLOETELVEPLTPSGAMPNQOMILKEQYI 720
 Db 661 ILLVVLGVVFGILIKSRQOKIRYTRRSLLOETELVEPLTPSGAMPNQOMILKEQYI 720
 QY 721 KANSKFTGITEL-TVYKGIWIPGENYKIPVAIKVLRENTSPRANKEILDEAVVMAVGVS 779
 Db 721 KANSKFTGITEL-TVYKGIWIPGENYKIPVAIKVLRENTSPRANKEILDEAVVMAVGVS 779
 QY 780 PYVSRLLGICLTSTVQVLYTQDMPYGCILDHVRENRGRLSQDILLNCKQIAKMSYLEBV 839
 Db 780 PYVSRLLGICLTSTVQVLYTQDMPYGCILDHVRENRGRLSQDILLNCKQIAKMSYLEBV 839
 QY 840 RLVHRDLAARNVLVKSNHYKIDTDFGLARLLDIDETEHADGGKVIKMMALLESILRRRF 899
 Db 840 RLVHRDLAARNVLVKSNHYKIDTDFGLARLLDIDETEHADGGKVIKMMALLESILRRRF 899
 QY 900 THSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLLEKGERLPPICTIDVYIMVWKCM 959
 Db 900 THSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLLEKGERLPPICTIDVYIMVWKCM 959
 QY 960 MIDSECRPRFRELVSSESRMARDPQRFVVIQNEIDLGPASFLDSTFYRSLLEDDMDGLVD 1019

```

Db      960 MIDSECRPRFRELVESEFSMAKDPORFVVIQNEUDGPRASPLDSTFFYSLEDDMDGDVD 1019
Qy      1020 AEEYLVPOQGFPCPDPAAGAGMHHRRSSSTRSGGDLTLGLEPSEEBAPRSPLAPSE 1079
Db      1020 AEEYLVPOQGFPCPDPAAGAGMHHRRSSSTRSGGDLTLGLEPSEEBAPRSPLAPSE 1079
Qy      1080 GAGSDVFDGDLGGAAGKGLQSLPTHDPSPLOKYSDDPTVPLPSETDGVVAPLTCSPQPEY 1139
Db      1080 GAGSDVFDGDLGGAAGKGLQSLPTHDPSPLOKYSDDPTVPLPSETDGVVAPLTCSPQPEY 1139
Qy      1140 VNQDVRPQPPSPFRSGPLPAAPAGATLERAKTSLPGKNGVYKDVFAFGAVENTEYLTIP 1199
Db      1140 VNQDVRPQPPSPFRSGPLPAAPAGATLERAKTSLPGKNGVYKDVFAFGAVENTEYLTIP 1199
Qy      1200 QGGAAPQPHPPAFSPAFDNLVYMDQDPERGAPSTFKGTPTAENPEYLGIDVPV 1255
Db      1200 QGGAAPQPHPPAFSPAFDNLVYMDQDPERGAPSTFKGTPTAENPEYLGIDVPV 1255

```

RESULT 8

AAV84780
ID AAV84780 standard; Protein; 1255 AA.

AC AAV84780;

DT 08-AUG-2000 (first entry)

DE Amino acid sequence of the SPLICE erdb-2 receptor protein.

XX SPLICE erdb-2 receptor protein; cell transformation disorder; cancer;

KM tumor cell proliferation; tissue degeneration; arthropathy;

KM bone resorption; inflammatory disease; degenerative disorder;

KM wound healing.

XX Homo sapiens.

OS MO200020579-A1.

XX 13-APR-2000.

XX 01-OCT-1999; 99MO-CH00912.

XX 02-OCT-1998; 98US-0165192.

XX (UYMC-) UNIV MCMASTER.

XX Muller WJ, Siegel PM;

XX WPI, 2000-303768/26.

XX N-PSDB; AAA14812.

XX Nucleic acid encoding an erdb 2 receptor protein designated SPLICE

PT erdb-2, inhibitors of the protein are useful for treatment of cancer -

XX Claim 3; Fig 2; 60pp; English.

XX The present sequence represents a SPLICE erdb-2 receptor protein. The
 CC protein has an in-frame deletion of 16 amino acids, 2 of which are
 CC conserved cysteine residues, compared to the unspliced protein. The
 CC erdb-2 polynucleotide is used to construct probes for detecting
 CC disorders of cell transformation such as cancer. Antibodies to the
 CC protein may be used to detect SPLICE erdb-2 in a sample. Agents of
 CC (e.g. antisense oligonucleotides) which inhibit the expression of
 CC SPLICE erdb-2 are useful for reducing tumor cell proliferation and
 CC treating cancer. Substances which stimulate SPLICE erdb-2 are useful
 CC in which degeneration of tissue occurs, such as arthropathy, bone
 CC resorption, inflammatory diseases, degenerative disorders of the
 CC central nervous system and wound healing.

XX Sequence 1255 AA;

```

Query Match      97.4%; Score 6640; DB 21; Length 1255;
Best Local Similarity 97.5%; Pred. No. 0;
Matches 1225; Conservative 8; Mismatches 21; Indels 2; Gaps 2

Qy      1 MELALCRMGILLALLPGAASVQVCTGTDMLRLPASDETHLDMRLHYQCCQVQGNL 60
Db      1 MELALCRMGILLALLPGAASVQVCTGTDMLRLPASDETHLDMRLHYQCCQVQGNL 60
Qy      61 ELTLPNASTLSEFQDIQEVGVVLIANHOVQVPLQRIYRGQVQLEEDNVALVLNDG 120
Db      61 ELTLPNASTLSEFQDIQEVGVVLIANHOVQVPLQRIYRGQVQLEEDNVALVLNDG 120
Qy      61 ELTLPNASTLSEFQDIQEVGVVLIANHOVQVPLQRIYRGQVQLEEDNVALVLNDG 120
Db      61 ELTLPNASTLSEFQDIQEVGVVLIANHOVQVPLQRIYRGQVQLEEDNVALVLNDG 120
Qy      121 DPLNNTTPVTGASPGGRLRELQSLTEILKGGVLIQBNPOLCYQDTIIMKDIHKNOLA 180
Db      121 DPLNNTTPVTGASPGGRLRELQSLTEILKGGVLIQBNPOLCYQDTIIMKDIHKNOLA 180
Qy      121 DPLNNTTPVTGASPGGRLRELQSLTEILKGGVLIQBNPOLCYQDTIIMKDIHKNOLA 180
Db      121 DPLNNTTPVTGASPGGRLRELQSLTEILKGGVLIQBNPOLCYQDTIIMKDIHKNOLA 180
Qy      181 LTLIDTRSRACHPCSPMCKGSRQWSESEDDQSLTRYCAGGACARCGPLPTDCCHQC 240
Db      181 LTLIDTRSRACHPCSPMCKGSRQWSESEDDQSLTRYCAGGACARCGPLPTDCCHQC 240
Qy      181 LTLIDTRSRACHPCSPMCKGSRQWSESEDDQSLTRYCAGGACARCGPLPTDCCHQC 240
Db      181 LTLIDTRSRACHPCSPMCKGSRQWSESEDDQSLTRYCAGGACARCGPLPTDCCHQC 240
Qy      241 AAGCTGPKHSDCLALHFNHSGICELHCPALVTYNTDFESMPNBEHRTFGASCVTACP 300
Db      241 AAGCTGPKHSDCLALHFNHSGICELHCPALVTYNTDFESMPNBEHRTFGASCVTACP 300
Qy      301 YNYLSTVGSCTLVCPHLNQEVTAEFGTCRCKSPCARVTCYIGMOYIKANSKFTGIT 360
Db      301 YNYLSTVGSCTLVCPHLNQEVTAEFGTCRCKSPCARVTCYIGMOYIKANSKFTGIT 360
Qy      301 YNYLSTVGSCTLVCPHLNQEVTAEFGTCRCKSPCARVTCYIGMOYIKANSKFTGIT 360
Db      301 YNYLSTVGSCTLVCPHLNQEVTAEFGTCRCKSPCARVTCYIGMOYIKANSKFTGIT 360
Qy      361 ELSPAGCKTIFGSLAF.PESFDDPASNAP.POPBOLQVFEETLEITGVLYISAMPDLSLP 420
Db      361 ELSPAGCKTIFGSLAF.PESFDDPASNAP.POPBOLQVFEETLEITGVLYISAMPDLSLP 420
Qy      361 IGFAGCKTIFGSLAF.PESFDDPASNAP.POPBOLQVFEETLEITGVLYISAMPDLSLP 420
Db      361 IGFAGCKTIFGSLAF.PESFDDPASNAP.POPBOLQVFEETLEITGVLYISAMPDLSLP 420
Qy      421 DLSVFQNLQVIRGRILHNGAVSLT.QGLISWLGRLSRLGSLALIHNTHLCEVHTV 480
Db      421 DLSVFQNLQVIRGRILHNGAVSLT.QGLISWLGRLSRLGSLALIHNTHLCEVHTV 480
Qy      421 DLSVFQNLQVIRGRILHNGAVSLT.QGLISWLGRLSRLGSLALIHNTHLCEVHTV 480
Db      421 DLSVFQNLQVIRGRILHNGAVSLT.QGLISWLGRLSRLGSLALIHNTHLCEVHTV 480
Qy      481 PMDQLFNPHQALHTANPREDECEVEGSLAQCLCARHGMGPGTQCVNCSQPLRGQEC 540
Db      481 PMDQLFNPHQALHTANPREDECEVEGSLAQCLCARHGMGPGTQCVNCSQPLRGQEC 540
Qy      481 PMDQLFNPHQALHTANPREDECEVEGSLAQCLCARHGMGPGTQCVNCSQPLRGQEC 540
Db      481 PMDQLFNPHQALHTANPREDECEVEGSLAQCLCARHGMGPGTQCVNCSQPLRGQEC 540
Qy      541 VEECRVLQGLPREVYNAHRLCP.HRCPQONGSVTCFGEADQCVAAHYNDPFCVARC 600
Db      541 VEECRVLQGLPREVYNAHRLCP.HRCPQONGSVTCFGEADQCVAAHYNDPFCVARC 600
Qy      541 VEECRVLQGLPREVYNAHRLCP.HRCPQONGSVTCFGEADQCVAAHYNDPFCVARC 600
Db      541 VEECRVLQGLPREVYNAHRLCP.HRCPQONGSVTCFGEADQCVAAHYNDPFCVARC 600
Qy      601 PSQVAPDLSYMPIWKPFDEGACQPCPINCCHSVDDDDKGCAPORASPLTISIAVVG 660
Db      601 PSQVAPDLSYMPIWKPFDEGACQPCPINCCHSVDDDDKGCAPORASPLTISIAVVG 660
Qy      601 PSQVAPDLSYMPIWKPFDEGACQPCPINCCHSVDDDDKGCAPORASPLTISIAVVG 660
Db      601 PSQVAPDLSYMPIWKPFDEGACQPCPINCCHSVDDDDKGCAPORASPLTISIAVVG 660
Qy      661 ILVVVLGVVFGILIKRQOKIRKTYMRLLQETELVEPLPSGAMPNQAQWRIKQYI 720
Db      661 ILVVVLGVVFGILIKRQOKIRKTYMRLLQETELVEPLPSGAMPNQAQWRIKQYI 720
Qy      661 ILVVVLGVVFGILIKRQOKIRKTYMRLLQETELVEPLPSGAMPNQAQWRIKQYI 720
Db      661 ILVVVLGVVFGILIKRQOKIRKTYMRLLQETELVEPLPSGAMPNQAQWRIKQYI 720
Qy      721 KANSKFTIGTEL.TYKKGIMIPGGEVVK.PVAKIKLRENTSKAKETILDEAYVMAAGVS 779
Db      721 KANSKFTIGTEL.TYKKGIMIPGGEVVK.PVAKIKLRENTSKAKETILDEAYVMAAGVS 779
Qy      721 KANSKFTIGTEL.TYKKGIMIPGGEVVK.PVAKIKLRENTSKAKETILDEAYVMAAGVS 779
Db      721 KANSKFTIGTEL.TYKKGIMIPGGEVVK.PVAKIKLRENTSKAKETILDEAYVMAAGVS 779
Qy      780 PYVSRLLGICLTSTVOLVTLQMPYGCGLDHEENRGRLSQDILNMCMQIAKAGSYLEDV 839
Db      780 PYVSRLLGICLTSTVOLVTLQMPYGCGLDHEENRGRLSQDILNMCMQIAKAGSYLEDV 839
Qy      780 PYVSRLLGICLTSTVOLVTLQMPYGCGLDHEENRGRLSQDILNMCMQIAKAGSYLEDV 839
Db      780 PYVSRLLGICLTSTVOLVTLQMPYGCGLDHEENRGRLSQDILNMCMQIAKAGSYLEDV 839
Qy      840 RLYHRDLAANVLYVSEBNVKTIDFGLALLIDITETVHAOGKVPITMMLLESTLRFR 899
Db      840 RLYHRDLAANVLYVSEBNVKTIDFGLALLIDITETVHAOGKVPITMMLLESTLRFR 899
Qy      840 RLYHRDLAANVLYVSEBNVKTIDFGLALLIDITETVHAOGKVPITMMLLESTLRFR 899
Db      840 RLYHRDLAANVLYVSEBNVKTIDFGLALLIDITETVHAOGKVPITMMLLESTLRFR 899
Qy      900 THQSDVSYGVTVWELMTGAKPYDIPAREIPDLLEKGERLPQPPICITIDVYIMVWKW 959
Db      900 THQSDVSYGVTVWELMTGAKPYDIPAREIPDLLEKGERLPQPPICITIDVYIMVWKW 959
Qy      900 THQSDVSYGVTVWELMTGAKPYDIPAREIPDLLEKGERLPQPPICITIDVYIMVWKW 959
Db      900 THQSDVSYGVTVWELMTGAKPYDIPAREIPDLLEKGERLPQPPICITIDVYIMVWKW 959
Qy      960 MIDSECRPRFRELVESEFSMAKDPORFVVIQNEUDGPRASPLDSTFFYSLEDDMDGDVD 1019
Db      960 MIDSECRPRFRELVESEFSMAKDPORFVVIQNEUDGPRASPLDSTFFYSLEDDMDGDVD 1019
Qy      1020 AEEYLVPOQGFPCPDPAAGAGMHHRRSSSTRSGGDLTLGLEPSEEBAPRSPLAPSE 1079

```

Db 1020 AEEYLVPQGGFFCDDPAPGAGCMVHRRSSSTRSGGDLTLGLEPSEEARPRSLAPSE 1079

QY 1080 GAGSDVFDGDLGMGAAGKQLQSLPTHDSPLQRYSDPVLPLPSETDGVAPLTSPPQPEY 1139

Db 1080 GAGSDVFDGDLGMGAAGKQLQSLPTHDSPLQRYSDPVLPLPSETDGVAPLTSPPQPEY 1139

QY 1140 VAGPDVPRPQPSRPGSPLPAARPAATLERPKTSLPGKNGVVKDVFAPFGAVENTPEYLT 1199

Db 1140 VAGPDVPRPQPSRPGSPLPAARPAATLERPKTSLPGKNGVVKDVFAPFGAVENTPEYLT 1199

QY 1200 QCGAAPQPHPPAPSPAPFDNLVYWDQDPPEPGAPSPSTKGTPTAENPEYIGLDVAV 1255

Db 1200 QCGAAPQPHPPAPSPAPFDNLVYWDQDPPEPGAPSPSTKGTPTAENPEYIGLDVAV 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein, 1255 AA.

XX AAB85458;
AC AAB85458;
DT 25-SEP-2001 (first entry)
XX Human HER-2/neu protein.
DE Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX Homo sapiens.
OS
PN MO200153463-A2.
XX 26-JUL-2001.
PD 19-JAN-2001; 2001MO-US01850.
PF 21-JAN-2000; 2000US-0177545.
PR (CORI-) CORIXA CORP.
FA Cheever MA, Hand-Zimmermann S;
PI MPI: 2001-476112/51.
XX N-PSDB; AAH23392.
DR
XX New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
PT for treating or preventing cancer, e.g. breast cancer -
XX
PS Claim 2; Page 41-46; 49pp; English.

CC The invention provides an isolated antigen-presenting cell, which
CC expresses at least an immunogenic portion of a polypeptide that produces
CC an immune response to HER-2/neu protein. The antigen-presenting cells are
CC useful as vaccines for eliciting or enhancing an immune response to
CC HER-2/neu protein, particularly in treating or preventing malignancies in
CC which the HER-2/neu oncogene is associated. Specifically, these are
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC colon, lung or prostate cancers. The present sequence represents
CC the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
SQ Sequence 1255 AA;

Query Match 97.4%; Score 6640; DB 22; Length 1255;
Best Local Similarity 97.5%; Pred. No. 0;
Matches 1225; Conservative 8; Mismatches 21; Indels 2; Gaps 2;

QY 1 MELALCRWGLLALPPGAASVCTGTDKRLPASPETHLDMLRHLYQGQGVVQGNL 60
Db 1 MELALCRWGLLALPPGAASVCTGTDKRLPASPETHLDMLRHLYQGQGVVQGNL 60
QY 61 ETTYPTNASTSLFDIOBVGCVLIHNVQVPLDRLKIVKGTQLFEDNYALAVLDNG 120

Db 61 ETTYPTNASTSLFDIOBVGCVLIHNVQVPLDRLKIVKGTQLFEDNYALAVLDNG 120

QY 121 DELNNTVTVAASPGGLRELOLSLEILKGGVLIQSNPOLCYQDTLLMKDIFHKQNOLA 180

Db 121 DELNNTVTVAASPGGLRELOLSLEILKGGVLIQSNPOLCYQDTLLMKDIFHKQNOLA 180

QY 181 LTLIDTNSRACHPESPMCKSRGWSSESDCGLTFTVACGACRKGPLPTDCCHQEC 240

Db 181 LTLIDTNSRACHPESPMCKSRGWSSESDCGLTFTVACGACRKGPLPTDCCHQEC 240

QY 241 AAGCTPGHSDCLACLFHNSGICELCPALVTYNTDTPFSMPNBEGRYTFGASCVTAC 300

Db 241 AAGCTPGHSDCLACLFHNSGICELCPALVTYNTDTPFSMPNBEGRYTFGASCVTAC 300

QY 301 YNYLSTDVSGCTLVCPPLHNOEVTLEDGTQRECKSKCAEVCYGLGMEHREYAVTSAN 360

Db 301 YNYLSTDVSGCTLVCPPLHNOEVTLEDGTQRECKSKCAEVCYGLGMEHREYAVTSAN 360

QY 361 ELEFAGCKKIFGSLAFUPESFDPAANTAPLOEBOQVETLEITGYLYISAMPDLP 420

Db 361 IOEFAGCKKIFGSLAFUPESFDPAANTAPLOEBOQVETLEITGYLYISAMPDLP 420

QY 421 DLSVFQNTQVIRGRILHNGAYSLTLOGLISWGLRSLEIGSGLAIHHNTHLCFVHTV 480

Db 421 DLSVFQNTQVIRGRILHNGAYSLTLOGLISWGLRSLEIGSGLAIHHNTHLCFVHTV 480

QY 481 PWDOLFRRPHQALLHTARPEDEVGSGIACHQICAGHGWGPGTCVNCOSFLRQEC 540

Db 481 PWDOLFRRPHQALLHTARPEDEVGSGIACHQICAGHGWGPGTCVNCOSFLRQEC 540

QY 541 VEECRVLQGLPREVYNARHCLPCHRECOPOGNSVTCGPEADOCVACAHKADPPFCVAC 600

Db 541 VEECRVLQGLPREVYNARHCLPCHRECOPOGNSVTCGPEADOCVACAHKADPPFCVAC 600

QY 601 PSGVPLDLSYMPIMKFPDEBACOPPCINCTHSCVDLDDGCPABOPASPLTIVSAVVG 660

Db 601 PSGVPLDLSYMPIMKFPDEBACOPPCINCTHSCVDLDDGCPABOPASPLTIVSAVVG 660

QY 661 ILLVVLGVVFGLIIRKROOKIRKYTRRLLLOETLEVEPLTPSGAMPNOAMILKEQYI 720

Db 661 ILLVVLGVVFGLIIRKROOKIRKYTRRLLLOETLEVEPLTPSGAMPNOAMILKEQYI 720

QY 721 KANSKFIGITEL-TVYGIWIPDGENVKIPVAIKVIRENTSPANKELIDEAYVMAVGVS 779

Db 721 KANSKFIGITEL-TVYGIWIPDGENVKIPVAIKVIRENTSPANKELIDEAYVMAVGVS 779

QY 780 PYVSLRGLICLTSTVOLVTOUMPYGCILDVRENRGLSGODLLNMCQIAKMSYLEDV 839

Db 780 PYVSLRGLICLTSTVOLVTOUMPYGCILDVRENRGLSGODLLNMCQIAKMSYLEDV 839

QY 840 RLVRHDLAARNVLYKSPNHVKITDPGLARLLDIDETRYHADGKVPIMMALESIARRF 899

Db 840 RLVRHDLAARNVLYKSPNHVKITDPGLARLLDIDETRYHADGKVPIMMALESIARRF 899

QY 900 THQSDWASVYTTWELMTFGAKPYDGIAPABIPDLLEKSGRLPQPICTIDVYIMVKKW 959

Db 900 THQSDWASVYTTWELMTFGAKPYDGIAPABIPDLLEKSGRLPQPICTIDVYIMVKKW 959

QY 960 MIDSECRPRFRELVESESRMARDPQRFVIVQNEDLGASPLDSTFYSRLIEDDDMDGLVD 1019

Db 960 MIDSECRPRFRELVESESRMARDPQRFVIVQNEDLGASPLDSTFYSRLIEDDDMDGLVD 1019

QY 1020 AEEYLVPQGGFFCDDPAPGAGCMVHRRSSSTRSGGDLTLGLEPSEEARPRSLAPSE 1079

Db 1020 AEEYLVPQGGFFCDDPAPGAGCMVHRRSSSTRSGGDLTLGLEPSEEARPRSLAPSE 1079

QY 1080 GAGSDVFDGDLGMGAAGKQLQSLPTHDSPLQRYSDPVLPLPSETDGVAPLTSPPQPEY 1139

Db 1080 GAGSDVFDGDLGMGAAGKQLQSLPTHDSPLQRYSDPVLPLPSETDGVAPLTSPPQPEY 1139

QY 1140 VAGPDVPRPQPSRPGSPLPAARPAATLERAKTSLPGKNGVVKDVFAPFGAVENTPEYLT 1199

DB 1140 VNQPDVRFQPPSPREGGLPAPRPAAGATLEBPCKTLSPCKGNVXKDVAFGAVENPEYLTP 1199
 QY 1200 QGGAAPQHPHPAPSPAFDNLYYWDQDPPERGA PPSFGKSTPTAENPEYLGLDVPV 1255
 DB 1200 QGGAAPQHPHPAPSPAFDNLYYWDQDPPERGA PPSFGKSTPTAENPEYLGLDVPV 1255

RESULT 10
 AAG88267
 ID AAG88267 standard; Protein; 1255 AA.
 AC AAG88267;
 DT 11-SEP-2001 (first entry)
 DE HER2/neu amino acid sequence.
 KM Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
 KM immune response; vaccine; tumour; cancer; cytotoxic; immunostimulant;
 KM tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
 OS Homo sapiens.
 PN MO200141787-A1.
 PD 14-JUN-2001.
 PF 11-DEC-2000; 2000MO-US33591.
 PR 10-DEC-1999; 99US-0458299.
 PA (EPRIM-) EPIMUNE INC.
 PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
 PI Keogh E;
 DR WPI; 2001-374995/39.
 PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
 PT cellular immune responses for the prevention and treatment of cancer -
 PS Disclosure; Page 15; 199pp; English.

The present invention describes isolated prepared HER2/neu epitopes (I).
 CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
 CC culture in vitro and binds to a complex of an epitope (I), bound to a
 CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
 CC and a second epitope and the peptide is less than 50 contiguous amino
 CC acids that have 100% identity with a native peptide sequence of HER2/neu;
 CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
 CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
 CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytotoxic
 CC and immunostimulant activities, and can be used in vaccines. (I), (II)
 CC and (III) are useful for inducing cellular immune responses for the
 CC prevention and treatment of cancer. (I) and (II) are useful for
 CC monitoring or evaluating an immune response to a tumour-associated
 CC antigen when incubated with a T lymphocyte sample from a patient and
 CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
 CC based vaccines mean that immunosuppressive epitopes that may be present
 CC in whole antigens may be avoided. Selected epitopes may be combined to
 CC enhance immunogenicity. The possible pathological side effects caused by
 CC infectious agents or whole protein antigen is eliminated. The vaccine
 CC provides the ability to direct and focus an immune response to multiple
 CC selected antigens from the same pathogen. Epitope-based anti-tumour
 CC vaccines provides the opportunity to combine epitopes derived from
 CC multiple tumour-associated molecules addressing the problem of tumour-
 CC antigen variability and reducing the likelihood of tumour escape due to
 CC antigen loss. AAG88266 to AAG8912 represent amino acid sequences used in
 CC the exemplification of the present invention.

Sequence 1255 AA;
 Query Match 97.4%; Score 6640; DB 22; Length 1255;

Best Local Similarity 97.5%; Pred. No. 0;
 Matches 1225; Conservative 8; Mismatches 21; Indels 2; Gaps 2

QY 1 MELAALCRMGILLALLPFGAASITQVCTGDMKRLPASEETHLMDLRHLYGCGVQVGNL 60
 DB 1 MELAALCRMGILLALLPFGAASITQVCTGDMKRLPASEETHLMDLRHLYGCGVQVGNL 60
 QY 61 ELTYLPTNALSFLQDIOEVQGVYLLAHNQVQVPLQRIYRGVQLFEDNVALAVLDNG 120
 DB 61 ELTYLPTNALSFLQDIOEVQGVYLLAHNQVQVPLQRIYRGVQLFEDNVALAVLDNG 120
 QY 121 DPLNNTPTVTGASPGGLRELQSLTEILKGGVLTQRNPOLCYODITLWKDIFHNQQLA 180
 DB 121 DPLNNTPTVTGASPGGLRELQSLTEILKGGVLTQRNPOLCYODITLWKDIFHNQQLA 180
 QY 181 LTLIDNRSRACHPCSPCKGSGCWSESSDDQSLRTVCAGGACRCKPLPTDCCHEQC 240
 DB 181 LTLIDNRSRACHPCSPCKGSGCWSESSDDQSLRTVCAGGACRCKPLPTDCCHEQC 240
 QY 241 AAGCTGPKASDCLACHFNHSGICEIHCALVTYNTDTFESMNPFGRTFGASCVTACP 300
 DB 241 AAGCTGPKASDCLACHFNHSGICEIHCALVTYNTDTFESMNPFGRTFGASCVTACP 300
 QY 301 YNTLSTDVSGCTLVCPAHQEVTAEDGTQCEKSKPCARVCYGLMOYIKANSFKIGIT 360
 DB 301 YNTLSTDVSGCTLVCPAHQEVTAEDGTQCEKSKPCARVCYGLMEHREVRVTSAN 360
 QY 361 ELTFAGCKKIFGSLAPSPFGDDPASNAPLOPQLOVFEITLITGYLYISAMPDILP 420
 DB 361 IQEFAQCKKIFGSLAPSPFGDDPASNAPLOPQLOVFEITLITGYLYISAMPDILP 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSLLTQGISLWGLRSRLRSGSLALIHNTLFCFRTV 480
 DB 421 DLSVFQNLQVIRGRILHNGAYSLLTQGISLWGLRSRLRSGSLALIHNTLFCFRTV 480
 QY 481 PMQQLFRNHQALLHTANPPEDECCYEGELACHQLCARHCHWGPFTQCVCNCSQFLRGEC 540
 DB 481 PMQQLFRNHQALLHTANPPEDECCYEGELACHQLCARHCHWGPFTQCVCNCSQFLRGEC 540
 QY 541 VEECRVTLQGLPREYVNAARCLPCHHECOPQNGSVTCFPEADQCAACHYVDPFCVARC 600
 DB 541 VEECRVTLQGLPREYVNAARCLPCHHECOPQNGSVTCFPEADQCAACHYVDPFCVARC 600
 QY 601 PGGVKPDLSYPMIKRPDEBEGACQCPINCTHSCVDDLDKCCPAEORASPLTISAVAG 660
 DB 601 PGGVKPDLSYPMIKRPDEBEGACQCPINCTHSCVDDLDKCCPAEORASPLTISAVAG 660
 QY 661 ILLVVLGVVFGILLKRRQOKIRKXTMRLLQETELVEPLTPSGAMPQAOAMRIKETEL 720
 DB 661 ILLVVLGVVFGILLKRRQOKIRKXTMRLLQETELVEPLTPSGAMPQAOAMRIKETEL 720
 QY 721 KANSKFIGITEL-TVYKGIWIPDGENVAKPAIKVLRNTPSKANKEILDEAYVMAGVS 779
 DB 721 KANSKFIGITEL-TVYKGIWIPDGENVAKPAIKVLRNTPSKANKEILDEAYVMAGVS 779
 QY 780 EYVSSLILGICLSTVQVLTQMLPFGCLLDHYRENRGRIGSDDLNMCKQIAKGSYLEDV 839
 DB 780 EYVSSLILGICLSTVQVLTQMLPFGCLLDHYRENRGRIGSDDLNMCKQIAKGSYLEDV 839
 QY 840 RLVHRLDAAENVLVSPHNVKITPFGARLLDIDETEYHADGGVPIKMMALBELLRRF 899
 DB 840 RLVHRLDAAENVLVSPHNVKITPFGARLLDIDETEYHADGGVPIKMMALBELLRRF 899
 QY 900 THQSVMSYGVTTWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYMIWKW 959
 DB 900 THQSVMSYGVTTWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYMIWKW 959
 QY 960 MIDSECRPRELSESPRMAADQRFVIVONEDLGASPLDSIFRYSLLEDMDGDLVD 1019
 DB 960 MIDSECRPRELSESPRMAADQRFVIVONEDLGASPLDSIFRYSLLEDMDGDLVD 1019
 QY 1020 ABEYIVPQGFCCDPAAGAGVWHHRSSSTRSGGDDTLTGLEPSEEA PRSPAPSE 1079
 DB 1020 ABEYIVPQGFCCDPAAGAGVWHHRSSSTRSGGDDTLTGLEPSEEA PRSPAPSE 1079

Dh 1020 AEEYLVPQGGFFCCHDPAPGAGMWHRRHSSSTRSGGDLTLGLPSEBEAPRSLAPSE 1079
Qy 1080 GAGSDVFDGDLGMAAGKLSLPTHDSPLOKRYSEDPYLPSETTGVAAPLTGSPQPEY 1139
Db 1080 GAGSDVFDGDLGMAAGKLSLPTHDSPLOKRYSEDPYLPSETTGVAAPLTGSPQPEY 1139
Qy 1140 VNQPDVPRQPPSPREBGLPAPAPAGATLEBAKTLSPGKNGVWVDVAFGAVENPEYLTP 1199
Db 1140 VNQPDVPRQPPSPREBGLPAPAPAGATLEBAKTLSPGKNGVWVDVAFGAVENPEYLTP 1199
Qy 1200 QGGAAPQPHPPAPSPAFDNLVYWDOPFERGAPSTFKGTPTAENPEYLGLDVPY 1255
Db 1200 QGGAAPQPHPPAPSPAFDNLVYWDOPFERGAPSTFKGTPTAENPEYLGLDVPY 1255

RESULT 11

AAE24067
ID AAE24067 standard; Protein: 1255 AA.

AC AAE24067;

DT 23-SEP-2002 (first entry)

DE Human Her-2 protein.

KM Human, Her-2; epidermal growth factor receptor 2; infection; cancer;
KM hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW tumour; gene therapy; phosphorochioate backbone.

OS Homo sapiens.

PN W0200222636-A1.

PD 21-MAR-2002.

PF 12-SEP-2001; 2001MO-US28572.

PR 15-SEP-2000; 2000US-0663834.

PA (ISIS-) ISIS PHARM INC.

PI Bennet CF, Cowsett LM;

DR WPI; 2002-471192/50.

DR N-PSDB; AAD38904.

PT Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT Inflammation or to prevent infection in humans -

PS Example 13; Page 95-107; 116pp; English.

CC The invention relates to antisense compounds targeted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.

SQ Sequence 1255 AA;

Query Match 97.4%; Score 6640; DB 23; Length 1255;
Best local Similarity 97.5%; Pred. No. 0.
Matches 1225; Conservative 8; Mismatches 21; Indels 2; Gaps 2;

Qy 1 MEIAALCRWGLLIALIPGAASVCTGTDMKLRIPASPTHLDMLRRLYQGCGVVGQNL 60
Db 1 MEIAALCRWGLLIALIPGAASVCTGTDMKLRIPASPTHLDMLRRLYQGCGVVGQNL 60

Qy 61 ELTYLPTNASTFLQDIEQVGVYLIAHNOVQVPLQRLIRVGTOLPEDNYALAVLNG 120
Db 61 ELTYLPTNASTFLQDIEQVGVYLIAHNOVQVPLQRLIRVGTOLPEDNYALAVLNG 120
Qy 121 DPLNNTTPVYGAAPGGLRELQRLSLTEILKGGVLIQIRNQLCYQDITLMDKDFHKNNQLA 180
Db 121 DPLNNTTPVYGAAPGGLRELQRLSLTEILKGGVLIQIRNQLCYQDITLMDKDFHKNNQLA 180
Qy 181 LTLIDTNRASACHPGCPMKCKSGKMSSESDQSLTRYCAGGCARCKPRLTDCHEQC 240
Db 181 LTLIDTNRASACHPGCPMKCKSGKMSSESDQSLTRYCAGGCARCKPRLTDCHEQC 240
Qy 241 AAGCTGKHSKDCACAHFNHSGICEHICALTYNTDTESMPNPGRTFFASCYTACP 300
Db 241 AAGCTGKHSKDCACAHFNHSGICEHICALTYNTDTESMPNPGRTFFASCYTACP 300
Qy 301 YNYLSTDVSSCTLVCPILHNOEVTAEADGTORCEKSKPCARVCYGLGMOYIKANSKFIGIT 360
Db 301 YNYLSTDVSSCTLVCPILHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVAVTSAN 360
Qy 361 ELEFAGCKKIFGSLAFIPESFGDDPASNTAPLOPQLOVFETLEETGLYISAMPDSLIP 420
Db 361 IQEPAGCKKIFGSLAFIPESFGDDPASNTAPLOPQLOVFETLEETGLYISAMPDSLIP 420
Qy 421 DLSVFQNLQVIRGRILHNGAVSLTQGLISWLGRLSRELSGHALIHNNHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAVSLTQGLISWLGRLSRELSGHALIHNNHLCFVHTV 480
Qy 481 PMQPLFRNPHQALLHNTANRPEDECVGEGLAGQLCARHCWGPPTCCVNCQSLRQGC 540
Db 481 PMQPLFRNPHQALLHNTANRPEDECVGEGLAGQLCARHCWGPPTCCVNCQSLRQGC 540
Qy 541 VEECRVLQGLPREYVARHCLCPHEPCQPNQNSVTCFGEADQCVAAHYXDPPECVAVRC 600
Db 541 VEECRVLQGLPREYVARHCLCPHEPCQPNQNSVTCFGEADQCVAAHYXDPPECVAVRC 600
Qy 601 PSGVKEPDLSTMPWKFPDEEGACQPCPINCTHSQVLDLCKGCPABQASPLTIVSAVVG 660
Db 601 PSGVKEPDLSTMPWKFPDEEGACQPCPINCTHSQVLDLCKGCPABQASPLTIVSAVVG 660
Qy 661 ILVVLVGVVFGILIRROOKIRKXTMRLLQETLVEPLPSSGAMPQAOQRLIKEQYI 720
Db 661 ILVVLVGVVFGILIRROOKIRKXTMRLLQETLVEPLPSSGAMPQAOQRLIKEQYI 720
Qy 721 KANSKFIGITELTYVKGIMIPDGENVKIPVAIKYLRNTSPKANKELIDBAVYVAGVGS 779
Db 721 KANSKFIGITELTYVKGIMIPDGENVKIPVAIKYLRNTSPKANKELIDBAVYVAGVGS 779
Qy 780 PYSRLILGICLSTVQVLYQLMPYGCCLLDHRENGRUGSODLNMCMQIAKMSYLEDV 839
Db 780 PYSRLILGICLSTVQVLYQLMPYGCCLLDHRENGRUGSODLNMCMQIAKMSYLEDV 839
Qy 840 RLVHRDLAARNLVSPNHVKITTDGLARLIDIDETEYHADGKVPKIMALESTLRERF 899
Db 840 RLVHRDLAARNLVSPNHVKITTDGLARLIDIDETEYHADGKVPKIMALESTLRERF 899
Qy 900 THOSDWSVGVVWELMTFGAKRYGIPAREIPDILEGERSLPPPICTIDVYIMVKCW 959
Db 900 THOSDWSVGVVWELMTFGAKRYGIPAREIPDILEGERSLPPPICTIDVYIMVKCW 959
Qy 960 MIDSECRPFRELVSBSFMSMARDPQFVIVONEDIGPASPIDSTFYSLIEDDDMGVLVD 1019
Db 960 MIDSECRPFRELVSBSFMSMARDPQFVIVONEDIGPASPIDSTFYSLIEDDDMGVLVD 1019
Qy 1020 AEEYLVPQGGFFCCHDPAPGAGMWHRRHSSSTRSGGDLTLGLPSEBEAPRSLAPSE 1079
Db 1020 AEEYLVPQGGFFCCHDPAPGAGMWHRRHSSSTRSGGDLTLGLPSEBEAPRSLAPSE 1079
Qy 1080 GAGSDVFDGDLGMAAGKLSLPTHDSPLOKRYSEDPYLPSETTGVAAPLTGSPQPEY 1139
Db 1080 GAGSDVFDGDLGMAAGKLSLPTHDSPLOKRYSEDPYLPSETTGVAAPLTGSPQPEY 1139
Qy 1140 VNQPDVPRQPPSPREBGLPAPAPAGATLEBAKTLSPGKNGVWVDVAFGAVENPEYLTP 1199

Db 1140 VNQPDVAFPPSPREGFLPAAPAGATLEPPKTLSPGKGVAVDVAFGAAVENEYLTLP 1199
 QY 1200 QGGAAQPPHPPAFSPAFDNLVYWDQDPERGAPSPSTFKGPTAEINPEYLGLDVV 1255
 Db 1200 QGGAAQPPHPPAFSPAFDNLVYWDQDPERGAPSPSTFKGPTAEINPEYLGLDVV 1255

RESULT 12
 AAE20479
 ID AAE20479 standard; Protein; 1255 AA.
 XX
 AC AAE20479;
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE Human Her-2/neu protein.
 XX
 KM Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
 KM human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Region 1021..1030
 FT /note= "Naturally processed HLA-B44-restricted epitope"
 PN MO200214503-A2.
 XX
 PD 21-FEB-2002.
 XX
 PF 14-AUG-2001; 2001WO-US41733.
 XX
 PR 14-AUG-2000; 2000US-225152P.
 PR 28-SEP-2000; 2000US-236428P.
 PR 21-FEB-2001; 2001US-270520P.
 XX
 PA (CORI-) CORIXA CORP.
 PI Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
 PI Mcneill PD, Vedvick TS;
 PI
 DR MPI: 2002-280758/32.
 DR N-PSDB; AAD32743.
 PS
 PT Novel isolated Her-2/Neu polypeptide composition useful for therapy,
 PT prevention and diagnosis of cancer, preferably breast cancer -
 XX
 PS Disclosure; Page 114-117; 129pp; English.
 CC The invention relates to an isolated Her-2/Neu polypeptide composition
 CC effective for eliciting an immune response. The invention is useful for
 CC eliciting an immune response in a patient, where the patient is human
 CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
 CC The composition is useful for the therapy and diagnosis of cancer.
 CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
 CC and other compositions for the diagnosis, prevention and treatment of
 CC human malignancies, for stimulating and/or expanding T cells specific for
 CC Her-2/neu polypeptide and for inhibiting the development of cancer in a
 CC patient. The invention is useful for stimulating a T cell response in a
 CC human patient, as probe or primer for nucleic acid hybridization, to
 CC selectively form duplex molecules with complementary stretches of the
 CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
 CC length gene from a suitable library, and to direct expression of a
 CC polypeptide in appropriate host cells. The composition is useful in
 CC prophylactic or therapeutic applications and for the treatment of cancer,
 CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
 CC associated malignancies. The invention is useful in gene therapy. The
 CC present sequence is human Her-2/neu protein.
 SQ Sequence 1255 AA;

Query Match 97.4%; Score 6640; DB 23; Length 1255;

Best Local Similarity 97.5%; Pred. No. 0;
 Matches 1225; Conservative 8; Mismatches 21; Indels 2; Gaps 2

QY 1 MELAAICRMWLLALLPPGAASQVCTGDMKLRLPASEETHLDMLRHLYGCGVVOGWL 60
 Db 1 MELAAICRMWLLALLPPGAASQVCTGDMKLRLPASEETHLDMLRHLYGCGVVOGWL 60
 QY 61 ELTYLPTNASLFLQDIOEVGVYVLIHNVQVQVPAQRIRIVRGTLQFEDNALAVLNG 120
 Db 61 ELTYLPTNASLFLQDIOEVGVYVLIHNVQVQVPAQRIRIVRGTLQFEDNALAVLNG 120
 QY 121 DPLNNTTPYTGASPGGLRELQLBSLTELKGGVLIQRNPOLCYQDTILMKDIFHNNQLA 180
 Db 121 DPLNNTTPYTGASPGGLRELQLBSLTELKGGVLIQRNPOLCYQDTILMKDIFHNNQLA 180
 QY 181 LTIIDNRSRACHPCSPMGKSGRCESESDQSILRTVCAGGACRGPPLPTDCHEQC 240
 Db 181 LTIIDNRSRACHPCSPMGKSGRCESESDQSILRTVCAGGACRGPPLPTDCHEQC 240
 QY 241 AAGCTGPKASDCLACLFHNSGICEIHCPLATVYNTDFESMPNDEGRYTFGASCVTACP 300
 Db 241 AAGCTGPKASDCLACLFHNSGICEIHCPLATVYNTDFESMPNDEGRYTFGASCVTACP 300
 QY 301 YNLTSDVSSCTLVCPFLHNGEYTAEDGTQRCCKSPCARVCYGLGMOYIKANSKFIGIT 360
 Db 301 YNLTSDVSSCTLVCPFLHNGEYTAEDGTQRCCKSPCARVCYGLGMOYIKANSKFIGIT 360
 QY 361 ELTEFAGCKKIFGSLAFLESPFDGDSANTAPLPQPOLQVFTEETGYLYISAMPDLSL 420
 Db 361 IQEFAQCKKIFGSLAFLESPFDGDSANTAPLPQPOLQVFTEETGYLYISAMPDLSL 420
 QY 421 DLSVFQNLQVIRGILHNGAYSLLTQGLGISWLGIRSLRELGSGLALIHNTHTLCFVATV 480
 Db 421 DLSVFQNLQVIRGILHNGAYSLLTQGLGISWLGIRSLRELGSGLALIHNTHTLCFVATV 480
 QY 481 PWDQFLRNPHOALLHTANRPEDECEGELACHQLCARHCGSPGTQCVNCSQFLRGEC 540
 Db 481 PWDQFLRNPHOALLHTANRPEDECEGELACHQLCARHCGSPGTQCVNCSQFLRGEC 540
 QY 541 VEECRVLOGLPREYVNAHCLPCHCECOPONGSVTCFPEADOCVACAHYDPPFCVARC 600
 Db 541 VEECRVLOGLPREYVNAHCLPCHCECOPONGSVTCFPEADOCVACAHYDPPFCVARC 600
 QY 601 PSGVCPDLSYMPIMKFPDDEBAGCQPCPINCCHSCVDLDDKCPAEPASPLTSIVSAVG 660
 Db 601 PSGVCPDLSYMPIMKFPDDEBAGCQPCPINCCHSCVDLDDKCPAEPASPLTSIVSAVG 660
 QY 661 ILVVVLGVFGIILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNOQMRILKEQYI 720
 Db 661 ILVVVLGVFGIILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNOQMRILKEQYI 720
 QY 721 KANSKFIGITEL-TYKGIWIPDENYKIPVAIKVLENTSPKANKETLDEAYVMAGVGS 779
 Db 721 KANSKFIGITEL-TYKGIWIPDENYKIPVAIKVLENTSPKANKETLDEAYVMAGVGS 779
 QY 780 PVSLLIICLTSTVQVLTQMPYGCGLDHYRENRKGLSGSOLLNMCQIAKMSYLEDV 839
 Db 780 PVSLLIICLTSTVQVLTQMPYGCGLDHYRENRKGLSGSOLLNMCQIAKMSYLEDV 839
 QY 840 RLVRDLAARNVLKSPNHYKITDFGLARLLIDETVYHAGDGVPIPKMAALSLIRRRF 899
 Db 840 RLVRDLAARNVLKSPNHYKITDFGLARLLIDETVYHAGDGVPIPKMAALSLIRRRF 899
 QY 900 THQSDWMSYGTVWELMTFGAKPYDGIIPAREIPDLLEKGRSLQPPICITIDVYMIWKW 959
 Db 900 THQSDWMSYGTVWELMTFGAKPYDGIIPAREIPDLLEKGRSLQPPICITIDVYMIWKW 959
 QY 960 MIDSECRPFRELVSEFSRMAKDPQRFVIVQNEIDLPASPLDSTFRSLIEDDMKDLDVD 1019
 Db 960 MIDSECRPFRELVSEFSRMAKDPQRFVIVQNEIDLPASPLDSTFRSLIEDDMKDLDVD 1019
 QY 1020 ABEYLVPQQGFCDDPAPGAGGVVHHRRSSSTRSGGDLTLGLPSEEEAPSPAPSE 1079
 Db 1020 ABEYLVPQQGFCDDPAPGAGGVVHHRRSSSTRSGGDLTLGLPSEEEAPSPAPSE 1079

Db 1020 AEEYLVPQGFPCDPAPAGAGVMVHRHSSSTRSGGDDLTGLERSEERAPSLAPSE 1079
 QY 1080 GAGSDVDPDGLGMAKAGLQSLTTHPSPLQRYSEPTVPLPSETDGYVAPLTCSPOPEY 1139
 Db 1080 GAGSDVDPDGLGMAKAGLQSLTTHPSPLQRYSEPTVPLPSETDGYVAPLTCSPOPEY 1139
 QY 1140 VNOPDVRFPPSPREGEPLPAARPAAGATLEBAKTLSSGKGVKVDYAFAGGAVENEYLTLP 1199
 Db 1140 VNOPDVRFPPSPREGEPLPAARPAAGATLEBAKTLSSGKGVKVDYAFAGGAVENEYLTLP 1199
 QY 1200 QGGAAAPQHPHPAPFSPAFDNLVWDODPPERGAPESTFKGPTAENPEYLGLDVEV 1255
 Db 1200 QGGAAAPQHPHPAPFSPAFDNLVWDODPPERGAPESTFKGPTAENPEYLGLDVEV 1255

RESULT 13
 AAM51143
 ID AAM51143 standard; Protein; 1255 AA.
 AC AAM51143;
 XX 17-JUN-2002 (first entry)
 DT
 XX Human Her-2/neu oncogene-encoded p185 glycoprotein.
 DE
 XX Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
 KM Tyrosine kinase; receptor; c-erbB2; gene therapy.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Domain 1..653 "extracellular domain"
 FT Domain 676..1255
 FT Domain 990..1255
 FT Domain /note= "intracellular domain"
 FT Domain /note= "phosphorylation domain"

PN MO200212341-A2.
 XX
 XX 14-FEB-2002.
 PD
 XX 03-AUG-2001; 2001MO-US24283.
 PF
 XX 03-AUG-2000; 2000US-0632507.
 PR
 XX (CORI-) CORIXA CORP.
 PA (SMK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX
 PI Cheever MA, Gheysen D;
 XX
 XX WPI: 2002-241743/29.
 DR N-PSDB; ABR92250.
 DR
 XX
 PT Her-2/neu fusion protein for treating or preventing cancer by eliciting
 PT or enhancing an immune response to the protein, has Her-2/neu
 PT extracellular domain fused to Her-2/neu intracellular or
 PT phosphorylation domain -
 PS
 XX Claim 68; Fig 7; 141pp; English.

CC The present sequence is that of human Her-2/neu (p185 glycoprotein
 CC or c-erbB2), an oncogenic self-protein and target for anti-cancer
 CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
 CC in a variety of cancers, including breast, ovarian, colon, lung and
 CC prostate cancer. Her-2/neu is a member of the tyrosine kinase
 CC family of receptor-like glycoproteins. It comprises an extracellular
 CC domain with homology to the epidermal growth factor receptor
 CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
 CC intracellular domain that also shows homology to EGFR. Its
 CC overexpression correlates with a poor prognosis in breast and
 CC ovarian cancers. The invention provides Her-2/neu fusion
 CC proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
 CC preferred fusion proteins, the extracellular domain of a Her-2/neu
 CC protein is fused to a Her-2/neu intracellular domain or
 CC phosphorylation domain (or its Delta fragment). An immune
 CC response to Her-2/neu protein is elicited or enhanced by
 CC administering the fusion protein in the form of a vaccine, or by
 CC transfecting cells of an animal *ex vivo* with a nucleic acid
 CC encoding the fusion protein, and delivering the transfected cells
 CC to the animal. The fusion proteins, nucleic acids, and isolated
 CC specific T-cells are useful for inhibiting the development of a
 CC cancer, especially breast, ovarian, colon, lung or prostate cancer
 CC in a patient. T cells that specifically react with a Her-2/neu
 CC fusion protein can be used to remove tumour cells from a sample in
 CC order to inhibit the development of cancer in a patient.
 XX
 SQ Sequence 1255 AA;
 Query Match 97.4%; Score 6640; DB 23; Length 1255;
 Best Local Similarity 97.5%; Pred. No. 0;
 Matches 1225; Conservative 8; Mismatches 21; Indels 2; Gaps 2;

QY 1 MELAALCRWGLLALLPQGAATQVCTGDMKRLPASPETHLMLRLYQCCVQGNL 60
 Db 1 MELAALCRWGLLALLPQGAATQVCTGDMKRLPASPETHLMLRLYQCCVQGNL 60
 QY 61 ELTYLPTNASLFLDIOEVQCYVLIANQVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLFLDIOEVQCYVLIANQVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120
 QY 121 DPLNNTTPVTGASPGGLRELDLRLSTLTLKGVLIQRNPQLCYODTLMKDTLHKNNOLA 180
 Db 121 DPLNNTTPVTGASPGGLRELDLRLSTLTLKGVLIQRNPQLCYODTLMKDTLHKNNOLA 180
 QY 181 LTLIDTNSRACHPCSPKCKSRGMBESSEDCQSLRTVACGACARCKGLPTDCHEQC 240
 Db 181 LTLIDTNSRACHPCSPKCKSRGMBESSEDCQSLRTVACGACARCKGLPTDCHEQC 240
 QY 241 AAGCTGPHNSDCLGLHNHSGIEMLCPALVYNTTFFSMPBERYTFGASCTTACP 300
 Db 241 AAGCTGPHNSDCLGLHNHSGIEMLCPALVYNTTFFSMPBERYTFGASCTTACP 300
 QY 301 YNYLSTDVSGCTLVCPHNGEVTAEEDTGCKECSKCAVCTGLGMEHLREYRAVTSAN 360
 Db 301 YNYLSTDVSGCTLVCPHNGEVTAEEDTGCKECSKCAVCTGLGMEHLREYRAVTSAN 360
 QY 361 ELEFAGCKTIFGSLAFLPESFDGPPASNTAPLOEQVFEETLEITGYLYISAMPDLP 420
 Db 361 ELEFAGCKTIFGSLAFLPESFDGPPASNTAPLOEQVFEETLEITGYLYISAMPDLP 420
 QY 421 DLSVFONLQVIRGRILINNGAYSLTLOGLISWGLRSLRGLALIHNTLTCFVHTV 480
 Db 421 DLSVFONLQVIRGRILINNGAYSLTLOGLISWGLRSLRGLALIHNTLTCFVHTV 480
 QY 481 PWDOLFNRPHQALHTANRPDECEVGEGLACHQLCARGHMGCPPTQVNCQFLRGQEC 540
 Db 481 PWDOLFNRPHQALHTANRPDECEVGEGLACHQLCARGHMGCPPTQVNCQFLRGQEC 540
 QY 541 VEECRVLOGLPREVYVNAHCLPCHPEQOPNGSVTFGPADQCVACAHYKBPFCVANC 600
 Db 541 VEECRVLOGLPREVYVNAHCLPCHPEQOPNGSVTFGPADQCVACAHYKBPFCVANC 600
 QY 601 PSQVPLSTWPIWKFDEBGAQCPPICTHSCVDLDKGCAPAEQRAPLTSIVSAVVG 660
 Db 601 PSQVPLSTWPIWKFDEBGAQCPPICTHSCVDLDKGCAPAEQRAPLTSIVSAVVG 660
 QY 661 ILVVLVGVVFGILIRKROOKIKYIMRLDTELVLP7TSGAMPNQAQRILKEQYI 720
 Db 661 ILVVLVGVVFGILIRKROOKIKYIMRLDTELVLP7TSGAMPNQAQRILKEQYI 720
 QY 721 KANSKFLGTEL-TYVKGIIWIPGDEVNKIPVAIKVLRNTSPKANKEIDEXYVNAAGVS 779
 Db 721 R-KVKVLGSGAFGTYYKGIWIPGDEVNKIPVAIKVLRNTSPKANKEIDEXYVNAAGVS 779

QY 780 PYVSRLLGICLTSTVOLVTOQMPYGCILDHVRENRRGLSODLLNMCQIAKMSYLEDV 839
 Db 780 PYVSRLLGICLTSTVOLVTOQMPYGCILDHVRENRRGLSODLLNMCQIAKMSYLEDV 839
 QY 840 RLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGKVP1KMALESILRRRF 899
 Db 840 RLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGKVP1KMALESILRRRF 899
 QY 900 THQSDVWSYGVTVELMTFGAKPYDGI PAEITDULEKGBRLPQPICTIDVYMMVNCW 959
 Db 900 THQSDVWSYGVTVELMTFGAKPYDGI PAEITDULEKGBRLPQPICTIDVYMMVNCW 959
 QY 960 MIDSECRPRELVESESRMARDPQRFVVIQNEDELGPASPLDSTFYRSLLDEDDMGDLVD 1019
 Db 960 MIDSECRPRELVESESRMARDPQRFVVIQNEDELGPASPLDSTFYRSLLDEDDMGDLVD 1019
 QY 1020 ABEYLYVPOQGFCDPAPGAGCMVHRHRSSTRSGGDLTLGLESESEARSLASE 1079
 Db 1020 ABEYLYVPOQGFCDPAPGAGCMVHRHRSSTRSGGDLTLGLESESEARSLASE 1079
 QY 1080 GAGSDVDFDGLGMAAGKQSLPPTHDSPLQRYSEDPTVPLPSETDGYVAPLTCSPQREY 1139
 Db 1080 GAGSDVDFDGLGMAAGKQSLPPTHDSPLQRYSEDPTVPLPSETDGYVAPLTCSPQREY 1139
 QY 1140 VNQPDVRRQPSREGEPLPARPAGATLERAKTLSPGKNGVVKDVAFAGAVENBEYLTP 1139
 Db 1140 VNQPDVRRQPSREGEPLPARPAGATLERAKTLSPGKNGVVKDVAFAGAVENBEYLTP 1139
 QY 1200 QCGAAPQHPHPAPSPAFDNLVYWDQDPPERGAPSPSTFGKPTAENBEYLGADV 1255
 Db 1200 QCGAAPQHPHPAPSPAFDNLVYWDQDPPERGAPSPSTFGKPTAENBEYLGADV 1255

RESULT 14
 AAU77114
 ID AAU77114 standard; Protein; 1255 AA.
 AC AAU77114;
 DT 05-JUN-2002 (first entry)
 DE Human Her-2/neu polypeptide.
 KW Human; Her-2/neu; cytosolic; haematological malignancy; CML; acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL; chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS; Hodgkin's lymphoma; T cell therapy.
 OS Homo sapiens.
 FN WO200213847-A2.
 PD 21-FEB-2002.
 PF 13-AUG-2001; 2001WO-US25408.
 PR 14-AUG-2000; 2000US-0638280.
 PS 28-SEP-2000; 2000US-0675904.
 PA (CORI-) CORIXA CORP.
 PI Gaiger A, Cheever MA, Hand-zimmermann S;
 DR WPI; 2002-280741/32.
 DB N-PSDB; ABK10730.
 XX Inhibiting haematological malignancy development by administering PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide encoding the polypeptide, or antigen presenting cells expressing the PT polypeptide -
 PS Disclosure; Page 71-74; 74pp; English.

CC The invention relates to a method for inhibiting development of CC haematological malignancy in a patient by administering a polypeptide CC comprising an immunogenic portion of Her-2/neu or a polynucleotide CC encoding the polypeptide. Antigen presenting cells that express the CC protein can also be administered. The sequences are used for inhibiting CC development of haematological malignancy such as acute myelogenous CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's CC lymphoma. This sequence represents the human Her-2/neu polypeptide.

Sequence 1255 AA;

Query Match 97.4%; Score 6640; DB 23; Length 1255;

Best Local Similarity 97.5%; Pred. No. 0;

Matches 1225; Conservative 8; Mismatches 21; Indels 2; Gaps 2

QY 1 MELAAICRKGILLALPPGAASTVCTGTDMKLRPASPETHLDMNLHYOGQVVOGSL 60
 Db 1 MELAAICRKGILLALPPGAASTVCTGTDMKLRPASPETHLDMNLHYOGQVVOGSL 60
 QY 61 ELTYLPTNASLFLQDIOEVQGYLIAHNVQVPLQRLIRVGTQDFEDNYALAVDNG 120
 Db 61 ELTYLPTNASLFLQDIOEVQGYLIAHNVQVPLQRLIRVGTQDFEDNYALAVDNG 120
 QY 121 DPLNNTPTVPGASPGGLRELOLRSTLEILKGVILIQNPOLCTQDTITLMDITFKNNQLA 180
 Db 121 DPLNNTPTVPGASPGGLRELOLRSTLEILKGVILIQNPOLCTQDTITLMDITFKNNQLA 180
 QY 181 LFLIDPTNSRAHPSPCKSKRCWGESSEDCOSLRTVACAGGARCKSLPDDCHEQC 240
 Db 181 LFLIDPTNSRAHPSPCKSKRCWGESSEDCOSLRTVACAGGARCKSLPDDCHEQC 240
 QY 241 AAGCTGPKXSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRYTFGASCYTACP 300
 Db 241 AAGCTGPKXSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRYTFGASCYTACP 300
 QY 301 YNYLSTDVQSCITVCPLENOEVTAEQDGTORCEKSKCAVCYGLGMQYKASKFSGIT 360
 Db 301 YNYLSTDVQSCITVCPLENOEVTAEQDGTORCEKSKCAVCYGLGMQYKASKFSGIT 360
 QY 361 ELEFAGCKKIFGSLAFLEBSFDGSPASNTABLOEQJVEFTLEITGYLYISAMPDSL 420
 Db 361 IOEFAGCKKIFGSLAFLEBSFDGSPASNTABLOEQJVEFTLEITGYLYISAMPDSL 420
 QY 421 DLSVQNLQVIRGRILHNGAYSLLTQGLGISWLSRLREUGSLALHNNHLCVHTV 480
 Db 421 DLSVQNLQVIRGRILHNGAYSLLTQGLGISWLSRLREUGSLALHNNHLCVHTV 480
 QY 481 PWDQLFKNPHQALLHTANRPEDECVGEGLACHQLCARHGMCGRPPTQCVNCSQFLRGQC 540
 Db 481 PWDQLFKNPHQALLHTANRPEDECVGEGLACHQLCARHGMCGRPPTQCVNCSQFLRGQC 540
 QY 541 VEECRVLOGPREYNAHCLPHEPCQPOWGSYTCGPRAQCVCAHAKDPFPCVAC 600
 Db 541 VEECRVLOGPREYNAHCLPHEPCQPOWGSYTCGPRAQCVCAHAKDPFPCVAC 600
 QY 601 PSQVAPDLSYPMKFPDEGACQPCFINTCHSCVDLDDKCGPAGQASPLTIIISAVVG 660
 Db 601 PSQVAPDLSYPMKFPDEGACQPCFINTCHSCVDLDDKCGPAGQASPLTIIISAVVG 660
 QY 661 ILLVYVAVGVGIIILKROQKIRYVTRRLLOETELVEPLTPSGAMPNQOMKILETEL 720
 Db 661 ILLVYVAVGVGIIILKROQKIRYVTRRLLOETELVEPLTPSGAMPNQOMKILETEL 720
 QY 721 KANSKFISITELTVYKGIWIPDGENVKIPVALIVLENTSPRANKELIDEAVYAVGVS 779
 Db 721 KANSKFISITELTVYKGIWIPDGENVKIPVALIVLENTSPRANKELIDEAVYAVGVS 779
 QY 780 PYVSRLLGICLTSTVOLVTOQMPYGCILDHVRENRRGLSODLLNMCQIAKMSYLEDV 839
 Db 780 PYVSRLLGICLTSTVOLVTOQMPYGCILDHVRENRRGLSODLLNMCQIAKMSYLEDV 839
 QY 840 RLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGKVP1KMALESILRRRF 899
 Db 840 RLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGKVP1KMALESILRRRF 899

```

Db      840 RLVHRLAARNVLYKSPNHVKITDGLARLDDIDETEVHADGGKVPKXMALESILRRRF 899
QY      900 THOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMWKCM 959
Db      900 THOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMWKCM 959
QY      960 MIDSECRPRFRELVSFGRMARDPQRFVVIQNEDLGASPPLDSTFYRSLLEDDMDGLVD 1019
Db      960 MIDSECRPRFRELVSFGRMARDPQRFVVIQNEDLGASPPLDSTFYRSLLEDDMDGLVD 1019
QY      1020 AEEYLVPOGGFFCPDPAPAGAGMWHHRSSSTSGGDDLTLGLEPSEEEAPRSLAPSE 1079
Db      1020 AEEYLVPOGGFFCPDPAPAGAGMWHHRSSSTSGGDDLTLGLEPSEEEAPRSLAPSE 1079
QY      1080 GAGSDVFDGDLGMAKGLQSLPTHDSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEY 1139
Db      1080 GAGSDVFDGDLGMAKGLQSLPTHDSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEY 1139
QY      1140 VNQDPVPRQPPSPRPGPIPARAPGATLERAKTILSPKNGVYKDVPAFGAVENPEYLTLP 1199
Db      1140 VNQDPVPRQPPSPRPGPIPARAPGATLERAKTILSPKNGVYKDVPAFGAVENPEYLTLP 1199
QY      1200 QGGAAPQPHPPAPSPADNLVYWDQDPPERGAPESTFKGTPTAENPEYLGIDVPY 1255
Db      1200 QGGAAPQPHPPAPSPADNLVYWDQDPPERGAPESTFKGTPTAENPEYLGIDVPY 1255

RESULT 15
AAR39568 standard; Protein; 1433 AA.
AC AAR39568;
AD 07-FEB-1994 (first entry)
DE Sequence of c-erbB-2 tumour antigen.
KM Tumour antigen; c-erbB-2; glycoprotein.
OS Homo sapiens.
PN M09316185-A.
PD 19-AUG-1993.
PF 05-FEB-1993; 93MO-US01055.
PR 06-FEB-1992; 92US-0831967.
XX (CETU ) CETUS ONCOLOGY CORP.
PA (CREA-) CREATIVE BIOMOLECULES INC.
PI Houston TX, Huston JS, Oppermann H, Ring DB;
XX WPI; 1993-272868/34.
DR N-PSDB; AAQ46083.
XX New single chain Fv polypeptide binding to C-erbB-2 tumour
PT antigen - for imaging or treating breast or ovarian cancer etc.
XX Disclosure; pages 48-54; 87pp; English.
PS
CC c-erbB-2 refers to a protein antigen expressed on the surface of
CC tumour cells, such as breast and ovarian tumour cells, which is an
CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39563 represents
CC the location of a stop codon in AAQ46083.
XX
SQ Sequence 1433 AA;

```

Query Match 96.8%; Score 6597; DB 14; Length 1433;
 Best Local Similarity 97.0%; Pred. No. 0;

```

Matches 1218; Conservative 10; Mismatches 26; Indels 2; Gaps 2
QY      1 MELALCRWGILLALPPGAASIQVCTGDMKRLRASPETHLDMRLHYQGVQVQNL 60
Db      1 MELALCRWGILLALPPGAASIQVCTGDMKRLRASPETHLDMRLHYQGVQVQNL 60
QY      61 ELTYLPTNASLSFQDIQEVQGYVLIAHNOVRVPLQRLIRVNGTOLFEDNYALVDNG 120
Db      61 ELTYLPTNASLSFQDIQEVQGYVLIAHNOVRVPLQRLIRVNGTOLFEDNYALVDNG 120
QY      121 DPLNNTTPVTGASPGGLREQLRSLTEILKGYLIQNPOLCYQDITLWKDIFKXNOLA 180
Db      121 DPLNNTTPVTGASPGGLREQLRSLTEILKGYLIQNPOLCYQDITLWKDIFKXNOLA 180
QY      181 LTLIDTNRSAHCSCSMCKSGSCSMGESSSDQSLRTYVCAAGCAACKPLPDDCCEOC 240
Db      181 LTLIDTNRSAHCSCSMCKSGSCSMGESSSDQSLRTYVCAAGCAACKPLPDDCCEOC 240
QY      241 AAGCTGKHSDDLACLHFNHSGICELHCPALVTYNTDFESMNPREGRYTFGASCTACP 300
Db      241 AAGCTGKHSDDLACLHFNHSGICELHCPALVTYNTDFESMNPREGRYTFGASCTACP 300
QY      301 YNYLSTVSGCTLYVCPILHNOEVTAEQTORCEKSCPCARVCTGLGQYIKANSKFIGIT 360
Db      301 YNYLSTVSGCTLYVCPILHNOEVTAEQTORCEKSCPCARVCTGLGQYIKANSKFIGIT 360
QY      361 ELFEAGCKKIFGSLAFLPESFDDPSANTAPLOPELOVETLEITGYLYISAMPDILP 420
Db      361 IQEPAGCKRIFGSLAFLPESFDDPSANTAPLOPELOVETLEITGYLYISAMPDILP 420
QY      421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRELSGIALIHNHNLHSFVHV 480
Db      421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRELSGIALIHNHNLHSFVHV 480
QY      481 PMDQLFENPHQALLHNPANREDCVCEGLACQLCARHGCMRPGPIQCCNGCQFLRGQEC 540
Db      481 PMDQLFENPHQALLHNPANREDCVCEGLACQLCARHGCMRPGPIQCCNGCQFLRGQEC 540
QY      541 VEEGRVLYQGLPREYVNASHCLEPHPCQCPQNSVTCFGEBAQCVAACHYKDPFCVARC 600
Db      541 VEEGRVLYQGLPREYVNASHCLEPHPCQCPQNSVTCFGEBAQCVAACHYKDPFCVARC 600
QY      601 PSQYKPDLSYMPIMKPPDEGACQPCPINCTHSQVLDLDCGCPAEGRASPLTISAVVG 660
Db      601 PSQYKPDLSYMPIMKPPDEGACQPCPINCTHSQVLDLDCGCPAEGRASPLTISAVVG 660
QY      661 ILVYVLYGVVFGILIKRQOKIRKYMRLQETELVEPLTSSGAMPQAOAMRLIKEQYI 720
Db      661 ILVYVLYGVVFGILIKRQOKIRKYMRLQETELVEPLTSSGAMPQAOAMRLIKEQYI 720
QY      721 KANSKFIGITEL-TYVKGIMIPDGENVKIPVAIKYLRENTSFRANKETILDEAYVAVGVS 779
Db      721 R-KVKYLGAGAFGTGVKGIWIPDGENVKIPVAIKYLRENTSFRANKETILDEAYVAVGVS 779
QY      780 PYSRLLGICLSTVQVLTQMLPYGCLDHYENGRIGSODLLMWCMQIAKGSYLEDV 839
Db      780 PYSRLLGICLSTVQVLTQMLPYGCLDHYENGRIGSODLLMWCMQIAKGSYLEDV 839
QY      840 RLVHRLAARNVLYKSPNHVKITDGLARLDDIDETEVHADGGKVPKXMALESILRRRF 899
Db      840 RLVHRLAARNVLYKSPNHVKITDGLARLDDIDETEVHADGGKVPKXMALESILRRRF 899
QY      900 THOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMWKCM 959
Db      900 THOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMWKCM 959
QY      960 MIDSECRPRFRELVSFGRMARDPQRFVVIQNEDLGASPPLDSTFYRSLLEDDMDGLVD 1019
Db      960 MIDSECRPRFRELVSFGRMARDPQRFVVIQNEDLGASPPLDSTFYRSLLEDDMDGLVD 1019
QY      1020 AEEYLVPOGGFFCPDPAPAGAGMWHHRSSSTSGGDDLTLGLEPSEEEAPRSLAPSE 1079
Db      1020 AEEYLVPOGGFFCPDPAPAGAGMWHHRSSSTSGGDDLTLGLEPSEEEAPRSLAPSE 1079

```

```
QY 1080 GAGSDVFDSDPLGMAAKGLQSLPTHDPSPLOQYSEDPYVPLPSETDGYVAPLTCSPQPEY 1139
Db 1080 GAGSDVFDSDPLGMAAKGLQSLPTHDPSPLOQYSEDPYVPLPSETDGYVAPLTCSPQPEY 1139
QY 1140 VNQPDVFRPQPPSPREGFLPAAPAPAGATLERAKTILSPGKNGVKDVFAGAVENPEYLTP 1199
Db 1140 VNQPDVFRPQPPSPREGFLPAAPAPAGATLERAKTILSPGKNGVKDVFAGAVENPEYLTP 1199
QY 1200 QGGAAPQPHPPPAFSAFDNLYWDDPPERGAAPSTFKGTPTAENPEYLGIDVPV 1255
Db 1200 QGGAAPQPHPPPAFSAFDNLYWDDPPERGAAPSTFKGTPTAENPEYLGIDVPV 1255
```

Search completed: July 22, 2003, 09:16:44
Job time : 42.9339 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:47:04 ; Search time 10.5741 Seconds

(without alignments)
4891.279 Million cell updates/sec

Title: SEQ4-653-675-12

Perfect score: 6776

Sequence: 1 MELALCRWGLLALLPFGA.....TFKQTPANPEYLGLDPEV 1247

Scoring table: BLOSUM62

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6602	97.4	1255	1	ERRB2_HUMAN
2	5815	85.8	1257	1	ERRB2_RAT
3	5799.5	85.6	1254	1	ERRB2_MESAU
4	3064	45.2	1210	1	EGFR_HUMAN
5	3033	44.8	1210	1	EGFR_MOUSE
6	2915.5	43.0	1308	1	ERRB4_HUMAN
7	2896	42.7	1308	1	ERRB4_RAT
8	2612.5	38.5	1167	1	XRMRK_XIPMA
9	2390.5	35.3	1342	1	ERRB3_HUMAN
10	2316.5	34.2	1339	1	ERRB3_RAT
11	1935	28.6	1426	1	EGFR_DROME
12	1666.5	24.6	634	1	ERRB1_ALV
13	1620	23.9	604	1	ERRB1_AVIER
14	1547	22.8	540	1	ERRB1_CHICK
15	1512	22.3	703	1	ERRB1_GALL
16	1287	19.0	1323	1	ERRB2_MOUSE
17	1142.5	16.9	245	1	ERRB2_MOUSE
18	725	10.8	1363	1	ERRB2_MOUSE
19	708	10.4	1300	1	ERRB2_MOUSE
20	704	10.4	1383	1	ERRB2_MOUSE
21	703.5	10.4	1372	1	ERRB2_MOUSE
22	699	10.3	1382	1	ERRB2_MOUSE
23	697	10.3	1297	1	ERRB2_MOUSE
24	696.5	10.3	1300	1	ERRB2_MOUSE
25	688	10.2	1607	1	ERRB2_MOUSE
26	682	10.1	1477	1	ERRB2_MOUSE
27	648	9.6	1367	1	ERRB2_MOUSE
28	635	9.4	1373	1	ERRB2_MOUSE
29	631.5	9.3	1370	1	ERRB2_MOUSE
30	618	9.1	1390	1	ERRB2_MOUSE
31	610	9.0	2146	1	ERRB2_MOUSE
32	593	8.8	1114	1	ERRB2_MOUSE
33	591	8.7	987	1	ERRB2_MOUSE

34	587.5	8.7	984	1	ERRB1_CHICK	Q07494	gallus gall
35	580.5	8.6	984	1	ERRB1_RAT	P09759	rattus norv
36	574.5	8.5	984	1	ERRB1_HUMAN	P54762	homo sapien
37	574	8.5	987	1	ERRB1_MOUSE	P54761	mus musc
38	571	8.4	757	1	ERRB1_HYDAT	P53356	hydra atten
39	571	8.4	902	1	ERRB1_XENLA	Q91736	xenopus lae
40	569	8.4	1068	1	ERRB1_XENLA	Q00944	gallus gall
41	566.5	8.4	1053	1	ERRB1_CHICK	Q03145	mus musc
42	565.5	8.3	977	1	ERRB1_MOUSE	P34152	rattus norv
43	563	8.3	1052	1	ERRB1_RAT	Q35346	rattus norv
44	561	8.3	1055	1	ERRB1_XENLA	Q91571	xenopus lae
45	560.5	8.3	985	1	ERRB1_XENLA		

ALIGNMENTS

RESULT 1
ERRB2_HUMAN STANDARD; PRT; 1255 AA.
AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE surface receptor HER2) (MLN 19).
GN ERBB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=66118663; PubMed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA Saito T., Toyoshima K.,
RT "Similarity of protein encoded by the human c-erbB-2 gene to
RT epidermal growth factor receptor.",
RL Nature 319:230-234(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=66070181; PubMed=2999974;
RA Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,
RA Francke U., Levinson A., Ulrich A.,
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene.",
RL Science 230:1132-1139(1985).
RN [3]
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=66016729; PubMed=2995967;
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.,
RA "A c-erbB-related protooncogene, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor receptor gene and is amplified in a
RT human salivary gland adenocarcinoma.",
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN [4]
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.,
RT "Characterization of a new allele of the human ERBB2 gene by allele-
RT specific competition hybridization.",
RL Genomics 15:426-429(1993).
RN [5]
RP FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
RN ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
RN ALPHA AND AMPHIREGULIN.
RN CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
RN tyrosine phosphate.
RN SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
RN (POTENTIAL).
RN SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- P1M: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES (BY SIMILARITY).
 CC POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN
 CC POSITIONS 654 AND 655. ALLELE B1 (654-TLE-TLE-655) HAS A FREQUENCY
 CC OF 0.783; ALLELE B2 (654-TLE-VAL-655) HAS A FREQUENCY OF 0.206;
 CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M11767; AAA35808.1; -
 DR EMBL: M11761; AAA35808.1; JOINED.
 DR EMBL: M11762; AAA35808.1; JOINED.
 DR EMBL: M11763; AAA35808.1; JOINED.
 DR EMBL: M11764; AAA35808.1; JOINED.
 DR EMBL: M11765; AAA35808.1; JOINED.
 DR EMBL: M11766; AAA35808.1; JOINED.
 DR EMBL: M11730; AAA75493.1; -
 DR EMBL: M12036; AAA35976.1; -
 DR EMBL: X03363; CAA27060.1; -
 DR PIR: A25491; A25491.
 DR PIR: A24571; A24571.
 DR HSSP: P1362; IFGK.
 DR Genew: HGNC:3430; ERBB2.
 DR MIM: 164870; -
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR Pfam: PF02757; YLP_2.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; FJ; 3.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Polymorphism.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
 FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 653 675 POTENTIAL.
 FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 720 987 PROTEIN KINASE.
 FT NP_BIND 726 734 ATP (BY SIMILARITY).
 FT BINDING 753 753 ATP (BY SIMILARITY).
 FT ACT_SITE 845 845 BY SIMILARITY.
 FT DISULFID 195 204 BY SIMILARITY.
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 220 227 BY SIMILARITY.
 FT DISULFID 224 235 BY SIMILARITY.
 FT DISULFID 236 244 BY SIMILARITY.
 FT DISULFID 240 252 BY SIMILARITY.
 FT DISULFID 255 264 BY SIMILARITY.
 FT DISULFID 268 295 BY SIMILARITY.
 FT DISULFID 299 311 BY SIMILARITY.
 FT DISULFID 315 331 BY SIMILARITY.
 FT DISULFID 334 338 BY SIMILARITY.
 FT DISULFID 511 520 BY SIMILARITY.
 FT DISULFID 515 528 BY SIMILARITY.
 FT DISULFID 531 540 BY SIMILARITY.
 FT DISULFID 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.
 FT DISULFID 567 584 BY SIMILARITY.
 FT DISULFID 587 596 BY SIMILARITY.
 FT DISULFID 600 623 BY SIMILARITY.
 FT DISULFID 626 634 BY SIMILARITY.
 FT DISULFID 630 642 BY SIMILARITY.
 FT MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 124 124 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 187 187 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 259 259 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 530 530 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 571 571 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARIANT 654 654 I->V.
 FT VARIANT 654 654 FTID=VAR_004077.
 FT VARIANT 655 655 I->V.
 FT CONFLICT 1170 1170 P->A (IN REF. 2).
 FT SEQUENCE 1255 AA; 137909 MM; 39BD5FD04DC962 CRC64;
 SQ

Query Match 97.4%; Score 6602; DB 1; Length 1255;
 Best Local Similarity 97.2%; Pred. No. 0;
 Matches 1220; Conservative 9; Mismatches 18; Indels 8; Gaps

QY 1 MELALACRWGLLALLPFGAASVCTGTDMKRLPASPEFTHLMLRHLVGGCCVVGONT 60
 DB 1 MELALACRWGLLALLPFGAASVCTGTDMKRLPASPEFTHLMLRHLVGGCCVVGONT 60
 QY 61 ELTYLPTNASLFLDIOEVGGVYLAHNOYVPLQSLRVRGTQFEDNYALAVLDNG 121
 DB 61 ELTYLPTNASLFLDIOEVGGVYLAHNOYVPLQSLRVRGTQFEDNYALAVLDNG 121
 QY 121 DELNNTPTPTGASPGGLRELOLRSTTEILKGGVLIQRNPOLCYODTILMKDIFHKNOLA 181
 DB 121 DELNNTPTPTGASPGGLRELOLRSTTEILKGGVLIQRNPOLCYODTILMKDIFHKNOLA 181
 QY 181 LTLIDNRSRACHPSPCKSRCKSGESSECCSLTTVCAGGACARCKGRLPDDCCHEQC 241
 DB 181 LTLIDNRSRACHPSPCKSRCKSGESSECCSLTTVCAGGACARCKGRLPDDCCHEQC 241
 QY 241 AAGCTGPRHSDCLACLFHNSGICELHCPALVTYNTDFESVPNPEGRYTFGASCVTACP 301
 DB 241 AAGCTGPRHSDCLACLFHNSGICELHCPALVTYNTDFESVPNPEGRYTFGASCVTACP 301
 QY 301 YNYLSTVGSCTLVCPHLNNOVTEADGTQRECKSKPCARVYCGLQKQYIKANSKFIGIT 361
 DB 301 YNYLSTVGSCTLVCPHLNNOVTEADGTQRECKSKPCARVYCGLQKQYIKANSKFIGIT 361
 QY 361 ELEFAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQVFEITLBEITGYLISAMPDULP 421
 DB 361 ELEFAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQVFEITLBEITGYLISAMPDULP 421
 QY 421 DLVSFQNLQVIRGILHNGAYSLTLQGLGISWGLRSLRELSSGALLIHNTLCEVHTV 481
 DB 421 DLVSFQNLQVIRGILHNGAYSLTLQGLGISWGLRSLRELSSGALLIHNTLCEVHTV 481
 QY 481 PMDQFRFPHALHTANRPDECVGEGALCHOLCARGHGCGGPPQCVNCSQFLMGQGC 541
 DB 481 PMDQFRFPHALHTANRPDECVGEGALCHOLCARGHGCGGPPQCVNCSQFLMGQGC 541
 QY 541 VEEGRVILQGLPREYVNAHCLPCHEPCQPNQSVTCFGEPAQCVACAHYKDPFPCVARG 601
 DB 541 VEEGRVILQGLPREYVNAHCLPCHEPCQPNQSVTCFGEPAQCVACAHYKDPFPCVARG 601
 QY 601 PSQVPLSTWPIKFPDEEGACQPCPINTHSCVLDKGCABRASPPLTISAVVG 661
 DB 601 PSQVPLSTWPIKFPDEEGACQPCPINTHSCVLDKGCABRASPPLTISAVVG 661
 QY 661 ILVVVVLGVVFGILLI---QYIKANS--KEIGITEL--PLTSGAMPNQAQRIIKETEL 712
 DB 661 ILVVVVLGVVFGILLI---QYIKANS--KEIGITEL--PLTSGAMPNQAQRIIKETEL 712
 QY 712 ILVVVVLGVVFGILLI---QYIKANS--KEIGITEL--PLTSGAMPNQAQRIIKETEL 720
 DB 712 ILVVVVLGVVFGILLI---QYIKANS--KEIGITEL--PLTSGAMPNQAQRIIKETEL 720


```

FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 626 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD_RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC...) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MM; 6129264583011402 CRC64;

```

```

Query Match 85.8%; Score 5815; DB 1; Length 1257;
Best Local Similarity 85.7%; Pred. No. 1,4e-104;
Matches 1077; Conservative 55; Mismatches 115; Indels 10; Gaps 5;

```

```

QY 1 MELAALCRWGLLALLPFGAASVQVCTGDMKLRUPASPEHLDMLRHLVQCCVYQGNL 60
D 1 MELAAMCRWGLLALLPFGIAGVQVCTGDMKLRUPASPEHLDMLRHLVQCCVYQGNL 60
QY 61 ELTYLPFNASTFLQIOEVQYVLIANNOVQVPLQRIATVKGQLFEDNYALAVLDNG 120
D 61 ELTYVPANASTFLQIOEVQYVLIANNOVQVPLQRIATVKGQLFEDNYALAVLDNR 120
QY 121 DPLNNTPTV-GASPGALRELOLRSLTEILKGVLIQRPOLCYODTILMKDIFKNNQL 179
D 121 DPODNVAASIPKTPGRLRELOLRSLTEILKGVLIQRPOLCYODTILMKDIFKNNQL 180
QY 180 ALTLIDTNSRAHPGSPMCKSGRCWSESDQCSTLRTVCAAGCCKPLPTDCHEQ 239
D 181 APVDIDTNSRAHPGSPMCKSGRCWSESDQCSTLRTVCAAGCCKPLPTDCHEQ 240
QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFEEMPERKRYFGASCYAC 299
D 241 CAGGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFEEMPERKRYFGASCYAC 300
QY 300 PNYVLTDSVGSCTLVCPPLHNOEVTADGTQRCCKSKPCARCYGLGMOYIKANSKFIGI 359
D 301 PNYVLTDSVGSCTLVCPPLHNOEVTADGTQRCCKSKPCARCYGLGMOYIKANSKFIGI 360
QY 360 TELEFAGCKKIFGSLAFLESFPGDPASTAPLOEOLQVFTLEITGYIYSWPNLSL 419
D 361 NVOEFGCKKIFGSLAFLESFPGDPASTAPLOEOLQVFTLEITGYIYSWPNLSL 420
QY 420 PDLVFQNLQVIRGRILHNGAYSLTLQGLISWLGRLSRLRELSGLALIHNTLCEVHT 479
D 421 PDLVFQNLQVIRGRILHNGAYSLTLQGLISWLGRLSRLRELSGLALIHNTLCEVHT 480
QY 480 VPWDOLFRRPHOHLHTANRDEDE-CVSGGLACHOGLCARHGWGPGPQCNCGQFLRQ 538
D 481 VPWDOLFRRPHOHLHTANRDEDE-CVSGGLACHOGLCARHGWGPGPQCNCGQFLRQ 540
QY 539 ECVEECRVLQGLPREYVNAHCLPCHPECOPONGSVTCFGPEADQCVCAHYKDPFCVA 598
D 541 ECVEECRVLQGLPREYVNAHCLPCHPECOPONGSVTCFGPEADQCVCAHYKDPFCVA 600
QY 599 RCSSGVKPDLSYPIWKPFDEGACQPCPINCCTHSCVDLDDKCPABORASPLTISVAV 658
D 601 RCSSGVKPDLSYPIWKPFDEGACQPCPINCCTHSCVDLDDKCPABORASPLTISVAV 660
QY 659 VGLILVVLGVVFGIIL---QYIKANS--KEIGTEL--PLTPSGAMPNOQRILKET 710
D 661 VGLILVVLGVVFGIIL---QYIKANS--KEIGTEL--PLTPSGAMPNOQRILKET 720
QY 711 ELRKVYVLSGAGFYVYKGIWIPDENYKIPVALIVLENTSPANKXEIIDEAYVMAVG 770
D 721 ELRKVYVLSGAGFYVYKGIWIPDENYKIPVALIVLENTSPANKXEIIDEAYVMAVG 780
QY 771 SPVSRLLIGICTLSTVQLVTQMLPFGCLDHRHNRGRLSGSDLLNMCQJAKMSYLED 830

```

```

D 781 SPVSRLLIGICTLSTVQLVTQMLPFGCLDHRHNRGRLSGSDLLNMCQJAKMSYLED 840
QY 831 VRLVRHDLAARVLYKSPNHVKTIDFGALRLDIDETEHADGKVPIKMALESILRR 890
D 841 VRLVRHDLAARVLYKSPNHVKTIDFGALRLDIDETEHADGKVPIKMALESILRR 900
QY 891 FTQSVMSVGYVWELMTFGKPPDGI PAREI PDLKSGEPLPPPICTI DVMIMWKC 950
D 901 FTQSVMSVGYVWELMTFGKPPDGI PAREI PDLKSGEPLPPPICTI DVMIMWKC 960
QY 951 WMIDSECRPRFELVSEFSRMARDQRFVIONEDLGPASPLDSTFTYSLEDDMDGLV 1010
D 961 WMIDSECRPRFELVSEFSRMARDQRFVIONEDLGPASPLDSTFTYSLEDDMDGLV 1020
QY 1011 DAEEYLVPOQGFCCDPDPAAGACGMMHRRSSSTSGGGLTLGLEPSEEPSPPLAPS 1070
D 1021 DAEEYLVPOQGFCCDPDPAAGACGMMHRRSSSTSGGGLTLGLEPSEEPSPPLAPS 1080
QY 1071 EGAGSDVFDGDLGMGAKGLQSLPHTDPSPLQRYSEDPTVLPSETDGVAPLTCSPPE 1130
D 1081 EGAGSDVFDGDLGMGAKGLQSLPHTDPSPLQRYSEDPTVLPSETDGVAPLTCSPPE 1140
QY 1131 YNQPDPVRPQPSPREGLPAPRPGATLERAKTSLSPKNGVYXDVFAFGAVENPEYLT 1190
D 1141 YNQPDPVRPQPSPREGLPAPRPGATLERAKTSLSPKNGVYXDVFAFGAVENPEYLT 1200
QY 1191 POGGAPOHPPPAPSPARFNNLYYNDODPBERGAPSPFXTPTAENEYGLDVPV 1247
D 1201 PREGTASPPHPSPAPSPARFNNLYYNDODPBERGAPSPFXTPTAENEYGLDVPV 1257

```

```

RESULT 3
ERR2_MESAU
ID ERR2_MESAU STANDARD; PRT: 1254 AA.
AC 060553;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
OS ERB2 OR NEU.
GN Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN 1;
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94193007; Pubmed=7908275;
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255(1994).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX.
CC ALTHOUGH NEUREGULIN DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; D16295; BAA03801.1; -
 CC HSSP; P11362; IFGK.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_Pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_Pkinase.
 DR InterPro; IPR004019; YLP_motif.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR Pfam; PF02757; YLP; 2.
 DR ProDom; PD000001; Euk_Pkinase; 1.
 DR SMART; SMO0261; FU; 3.
 DR SMART; SMO0219; TykC; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Proto-oncogene; Disease mutation.
 FT SIGNAL 1 21
 FT CHAIN 1 21
 FT DOMAIN 22 1254
 FT TRANSMEM 22 652
 FT DOMAIN 676 1254
 FT DOMAIN 158 368
 FT DOMAIN 472 644
 FT DOMAIN 720 987
 FT NP_BIND 726 734
 FT BINDING 753 753
 FT ACT_SITE 845 845
 FT DISULFID 195 204
 FT DISULFID 199 212
 FT DISULFID 236 244
 FT DISULFID 240 252
 FT DISULFID 255 264
 FT DISULFID 268 295
 FT DISULFID 299 311
 FT DISULFID 315 331
 FT DISULFID 334 338
 FT DISULFID 511 520
 FT DISULFID 515 528
 FT DISULFID 531 540
 FT DISULFID 544 560
 FT DISULFID 563 576
 FT DISULFID 567 584
 FT DISULFID 587 596
 FT DISULFID 600 623
 FT DISULFID 626 634
 FT DISULFID 630 642
 FT MOD_RES 1139 1139
 FT MOD_RES 1247 1247
 FT CARBOHYD 68 68
 FT CARBOHYD 125 125
 FT CARBOHYD 187 187
 FT CARBOHYD 259 259
 FT CARBOHYD 530 530
 FT CARBOHYD 571 571
 FT CARBOHYD 629 629
 FT VARIANT 658 658
 FT VARIANT 659 659
 SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

Query Match 85.6%; Score 5799.5; DB 1; Length 1254;
 Best Local Similarity 85.3%; Pred. No. 9-5e-304;
 Matches 1070; Conservative 64; Mismatches 112; Indels 9; Gaps 4;

QY 1 MELALCRMGLIALLPQAASTQVCTGTDMKRLPASPETHLDMRLHYGGCVVQGNL 60
 Db 1 MELAMCGMGLIALLPQAASTQVCTGTDMKRLPASPETHLDMRLHYGGCVVQGNL 60

QY ELTYLPNNAISLSPFQD:QEVGYVLLAHNOVROVPIORLRIYRGTOLEFEDNVALAVLQNG 120
 Db ELTYLPANATISLSPQD:QEVGYVLLAHNOVRAVPIORLRIYRGTOLEFEDKALAVLQNR 120
 QY DPLNNTTPVTGASFGRLRE:QLRSLTEILKGVLLGRNPOLCYQDTILMKDIFPKNNOLA 180
 Db DPLDNTTATGRTPEGLRE:QLRSLTEILKGVLLGRNPOLCYQDTILMKDIFPKNNOLA 180
 QY LTLIDTNRASRACPCSPMCKSGRCKWESSSDQSLRTYVACGACACKGPLPTDCHEOC 240
 Db PVLDITNRSPACPCACCKDNCKWASPEDCQTLTGITAPRAVPAARALPTDCHEOC 240
 QY AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTESMNPGRATFGASCYTAP 300
 Db AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTESMNPGRATFGASCYTAP 300
 QY YNLTSTVGSCTLVCPLNQEVTAEDGTORCEKSPCARVCYGLGMQYIKANSKFIGIT 360
 Db YNLTSTVGSCTLVCPLNQEVTAEDGTORCEKSPCARVCYGLGMHRLGARAITSAN 360
 QY ELBFGAGCKTIFGSLAF:PESEFDGDPASNTAPLOPELOVFEITLLETGYLISAMPDISL 420
 Db IOEFAGCKTIFGSLAF:PESEFDGDPASNTAPLOPELOVFEITLLETGYLISAMPDISL 420
 QY DLSVFQNLQYIRGRILHNGAYS:LTLOGLSIWLGRSLRELSGGLALIHNTLFCVHTV 480
 Db DLSVFQNLQYIRGRILHNGAYS:LTLOGLSIWLGRSLRELSGGLALIHNTLFCVHTV 480
 QY PMOQLPRNPQALLHRTANRPEDECEVEGLACIQLCARGHGMWGPPTQVNCQSPILRGEC 540
 Db PMOQLPRNPQALLHRTANRPEDECEVEGLACIQLCARGHGMWGPPTQVNCQSPILRGEC 540
 QY VBEGRVLOGLPREYVNAHRLCPHEPECQFQNSVTCFGEADQCAVAKYKDPRECVARC 600
 Db VBEGRVLOGLPREYVNAHRLCPHEPECQFQNSVTCFGEADQCAVAKYKDPRECVARC 600
 QY VKECRVWKGGPREYVNGKGLCPHEPECQFQNSVTCFGEADQCAVAKYKDPRECVARC 600
 Db VKECRVWKGGPREYVNGKGLCPHEPECQFQNSVTCFGEADQCAVAKYKDPRECVARC 600
 QY PSQVKNLSTMPYWKFPDEEGACQPCPINCSTSCVDLDDKCPABORASPLTISAVAG 660
 Db PSQVKNLSTMPYWKFPDEEGACQPCPINCSTSCVDLDDKCPABORASPLTISAVAG 660
 QY PSQVKNLSTMPYWKFPDEEGACQPCPINCSTSCVDLDDKCPABORASPLTISAVAG 660
 Db PSQVKNLSTMPYWKFPDEEGACQPCPINCSTSCVDLDDKCPABORASPLTISAVAG 660
 QY ILVWVVLGVVFGILL:---QYIKANS--KFIGITEL--PLTPSGAMPQAOIRLITKTEL 712
 Db ILVWVVLGVVFGILL:---QYIKANS--KFIGITEL--PLTPSGAMPQAOIRLITKTEL 712
 QY ILVWVVLGVVFGILL:---QYIKANS--KFIGITEL--PLTPSGAMPQAOIRLITKTEL 712
 Db ILVWVVLGVVFGILL:---QYIKANS--KFIGITEL--PLTPSGAMPQAOIRLITKTEL 712
 QY RKVKVLGSGAFGVYKGIWIPGENVVKIPVALKVLRENTSPKANKELIDEAVYVAGVSP 772
 Db RKVKVLGSGAFGVYKGIWIPGENVVKIPVALKVLRENTSPKANKELIDEAVYVAGVSP 772
 QY RKVKVLGSGAFGVYKGIWIPGENVVKIPVALKVLRENTSPKANKELIDEAVYVAGVSP 772
 Db RKVKVLGSGAFGVYKGIWIPGENVVKIPVALKVLRENTSPKANKELIDEAVYVAGVSP 772
 QY YVSRLLGICLTSTVQLVQLMRYGCLLHVMBENRGLSGODLLMWQVIAKMSYLEDVR 832
 Db YVSRLLGICLTSTVQLVQLMRYGCLLHVMBENRGLSGODLLMWQVIAKMSYLEDVR 832
 QY YVSRLLGICLTSTVQLVQLMRYGCLLHVMBENRGLSGODLLMWQVIAKMSYLEDVR 832
 Db YVSRLLGICLTSTVQLVQLMRYGCLLHVMBENRGLSGODLLMWQVIAKMSYLEDVR 832
 QY LVHRDLAANVLYKSPNHVKITDFGLARLLIDETEVYADGKVPKIMMALESLIRRF 892
 Db LVHRDLAANVLYKSPNHVKITDFGLARLLIDETEVYADGKVPKIMMALESLIRRF 892
 QY LVHRDLAANVLYKSPNHVKITDFGLARLLIDETEVYADGKVPKIMMALESLIRRF 892
 Db LVHRDLAANVLYKSPNHVKITDFGLARLLIDETEVYADGKVPKIMMALESLIRRF 892
 QY HOSDWVSGVTVWELMTGAKPYDGI:PAKEI:PDLLKEKERLPORPCTIDVYMIYVCMW 952
 Db HOSDWVSGVTVWELMTGAKPYDGI:PAKEI:PDLLKEKERLPORPCTIDVYMIYVCMW 952
 QY HOSDWVSGVTVWELMTGAKPYDGI:PAKEI:PDLLKEKERLPORPCTIDVYMIYVCMW 952
 Db HOSDWVSGVTVWELMTGAKPYDGI:PAKEI:PDLLKEKERLPORPCTIDVYMIYVCMW 952
 QY IDSECRPFRELVSSEFSEMARDDPQRFVYI:QNEDI:GPASPLDSTYRSLDEDDMDGLVDA 1012
 Db IDSECRPFRELVSSEFSEMARDDPQRFVYI:QNEDI:GPASPLDSTYRSLDEDDMDGLVDA 1012
 QY IDSECRPFRELVSSEFSEMARDDPQRFVYI:QNEDI:GPASPLDSTYRSLDEDDMDGLVDA 1012
 Db IDSECRPFRELVSSEFSEMARDDPQRFVYI:QNEDI:GPASPLDSTYRSLDEDDMDGLVDA 1012
 QY EBYLVPOQGFCCPDAPAGQMVHRRHSSTRSGGDLTLGLSPSESEAPRSLAPSEG 1072
 Db EBYLVPOQGFCCPDAPAGQMVHRRHSSTRSGGDLTLGLSPSESEAPRSLAPSEG 1072
 QY EBYLVPOQGFCCPDAPAGQMVHRRHSSTRSGGDLTLGLSPSESEAPRSLAPSEG 1072
 Db EBYLVPOQGFCCPDAPAGQMVHRRHSSTRSGGDLTLGLSPSESEAPRSLAPSEG 1072
 QY AGSDVFDLDMGAAGKIQSLPTHDPSPLQRYISDPTVLPJSEFDGYVADLTCSPOGEYV 1132
 Db AGSDVFDLDMGAAGKIQSLPTHDPSPLQRYISDPTVLPJSEFDGYVADLTCSPOGEYV 1132
 QY AGSDVFDLDMGAAGKIQSLPTHDPSPLQRYISDPTVLPJSEFDGYVADLTCSPOGEYV 1132
 Db AGSDVFDLDMGAAGKIQSLPTHDPSPLQRYISDPTVLPJSEFDGYVADLTCSPOGEYV 1132

```

Oy 1133 NOPDVPRPPSPREGPLPAARFAGATLERAKTLSPKGNVYKDVFAFGAVENPEYLTPO 1192
Db 1141 NQPEVPRPPPLTEGFLPVPYRPGATLERKTLSPKGNVYKDVFFGGAIVENPEYLVPR 1200
Oy 1193 GGAAPQPHPPAPSPADNLYYWDQDPPERGAPSPSTFKTPTTANPEYIGLDVPIV 1247
Db 1201 GGSASQPH-PPALCPAFDNLNLYWDQDPPERGSPSPNTFEGTPTAENPEYIGLDVPIV 1254

RESULT 4
EGFR_HUMAN STANDARD; PRT: 1210 AA.
AC P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; Q00732;
AC Q00688; Q9B252; Q9H2C9; Q9H2C9; Q9H3C9;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (BC 2.7.1.112) (Receptor
DE protein-tyrosine kinase ErbB-1).
GN EGFR OR ERBB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=84219729; PubMed=6328312;
RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
RT "Human epidermal growth factor receptor cDNA sequence and aberrant
RT expression of the amplified gene in A431 epidermoid carcinoma cells.";
RL Nature 309:418-425(1984).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX TISSUE=Placenta;
RX MEDLINE=95382957; PubMed=7654368;
RA Ilekis J.V., Stark B.C., Scoccia B.;
RT "Possible role of variant RNA transcripts in the regulation of
RT epidermal growth factor receptor expression in human placenta.";
RL Mol. Reprod. Dev. 41:149-156(1995).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX TISSUE=Placenta;
RX MEDLINE=97078686; PubMed=8918811;
RA Reiter J.L., Maibale N.J.;
RT "A 1.8 kb alternative transcript from the human epidermal growth
RT factor receptor gene encodes a truncated form of the receptor.";
RL Nucleic Acids Res. 24:4050-4056(1996).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX TISSUE=Placenta;
RX MEDLINE=97256547; PubMed=9103388;
RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
RT "Expression of a truncated epidermal growth factor receptor-like
RT protein (TEGFR) in ovarian cancer.";
RL Gynecol. Oncol. 65:36-41(1997).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
RX TISSUE=Placenta;
RX MEDLINE=21100872; PubMed=1161793;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Schein, Sinclair C., Pearisall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramanian S., Crossley T.D., Magnuson T.R., James C.D.,
RA Maibale N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative EGFR transcripts encoding truncated receptor
RT isoforms.";
RL Genomics 71:1-20(2001).
RN [6]
RP SEQUENCE OF 575-687 FROM N.A.
RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schein C.M.,
RA Lampland A.L., Balasubramanian S., Crossley T.O., Magnuson T.R.,

```

```

RA Maibale N.J.;
RT "Human and mouse alternative EGFR transcripts encoding only the
RT extracellular domain of the receptor.";
RL Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.
RN [7]
RP SEQUENCE OF 713-924 FROM N.A.
RX MEDLINE=84196372; PubMed=6326261;
RA Lin C.R., Chen W.S., Krueger W., Stolarsky L.S., Weber W.,
RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
RT "Expression cloning of human EGF receptor complementary DNA: gene
RT amplification and three related messenger RNA products in A431
RT cells.";
RL Science 224:843-848(1984).
RN [8]
RP SEQUENCE OF 150-962 FROM N.A.
RX MEDLINE=84245835; PubMed=6330563;
RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
RA Roe B.A., Merlino G.T., Pastan I.;
RT "Human epidermal growth factor receptor cDNA is homologous to a
RT variety of RNAs overproduced in A431 carcinoma cells.";
RL Nature 309:806-810(1984).
RN [9]
RP SEQUENCE OF 1028-1210 FROM N.A.
RX MEDLINE=85046483; PubMed=6093780;
RA Stamen F.A., Gope M.L., Schultz T.Z., Wright D.A., Carpenter G.,
RA O'Walley B.W.;
RT "Isolation of an evolutionarily conserved epidermal growth factor
RT receptor cDNA from human A431 carcinoma cells.";
RL Biochem. Biophys. Res. Commun. 124:1125-112(1984).
RN [10]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=88217333; PubMed=3329716;
RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
RA Waterfield M.D.;
RT "The human EGF receptor gene: structure of the 110 kb locus and
RT identification of sequences regulating its transcription.";
RL Oncogene Res. 1:375-386(1987).
RN [11]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=91107677; PubMed=1988448;
RA Haley J.D., Waterfield M.D.;
RT "Contributory effects of de novo transcription and premature
RT transcript termination in the regulation of human epidermal growth
RT factor receptor proto-oncogene RNA synthesis.";
RL J. Biol. Chem. 266:1746-1753(1991).
RN [12]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=85270438; PubMed=2991899;
RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
RT "Characterization and sequence of the promoter region of the human
RT epidermal growth factor receptor gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
RN [13]
RP SEQUENCE OF 540.
RA Kohda D.;
RL Submitted (SEP-1997) to the SWISS-PROT data bank.
RN [14]
RP RECEPTOR ACTIVITY.
RX MEDLINE=84191554; PubMed=6325948;
RA Mroczkowski B., Mosig G., Cohen S.;
RT "ATP-stimulated interaction between epidermal growth factor receptor
RT and supercoiled DNA.";
RL Nature 309:270-273(1984).
RN [15]
RP PHOSPHORYLATION.
RX MEDLINE=89278137; PubMed=2543678;
RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,
RA Howk R., Glyvol D., Ullrich A., Schlessinger J.;
RT "All autophosphorylation sites of epidermal growth factor (EGF)
RT receptor and HER2/neu are located in their carboxyl-terminal tails.
RT Identification of a novel site in EGF receptor.";
RL J. Biol. Chem. 268:10667-10671(1989).

```


QY 1136 DYPQPSRPSRGPAPARPAATLERAKTISPGKNGVAKVAFGAVENPEYL-TPQCG 1194
 Db 1096 SVPRFPAAGSVQNPVYHQNLP-----APSRDPHYQD--PHSTAQNPETLNVQ-- 1143
 QY 1195 AARPPHPPAPSPAFNDLYWDQ-----DP-----PRGAPPSFTFKTPTAENP 1238
 Db 1144 -----PTCVNSTFDSPPAHMAQGHSHQISLNDPDYQDFFPEAKRPNQIFKGS-TAENA 1195
 QY 1239 EYL 1241
 Db 1196 EYL 1198
 RESULT 5
 EGRF_MOUSE STANDARD; PRT; 1210 AA.
 ID EGRF_MOUSE 001279;
 AC 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (BC 2.7.1.112).
 GN EGRF.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Liver;
 RX MEDLINE=93026370; PubMed=1408137;
 RA Avivi A., Skorecki K., Yayon A., Givol D.;
 RT "Promoter region of the murine fibroblast growth factor receptor 2
 RT (bek/KGFR) gene.";
 RL Oncogene 7:1957-1962(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; and CD-1; TISSUE=Liver, and Decidua;
 RX MEDLINE=93126380; PubMed=678348;
 RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
 RT "Expression of the epidermal growth factor receptor gene is regulated
 RT in mouse blastocysts during delayed implantation.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Liver;
 RA Hibbs M.L.;
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B6/C3; TISSUE=Liver;
 RX MEDLINE=94170986; PubMed=8125255;
 RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
 RA Jenkins N.A., Lee D.C.;
 RT "The mouse waved-2 phenotype results from a point mutation in the EGF
 RT receptor tyrosine kinase.";
 RL Genes Dev. 8:399-413(1994).
 RN [5]
 RP SEQUENCE OF 1-714 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=91232866; PubMed=2030916;
 RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;
 RT "Comparison of EGF receptor sequences as a guide to study the ligand
 RT binding site.";
 RL Oncogene 6:673-676(1991).
 RN [6]
 RP SEQUENCE OF 969-1117 FROM N.A.
 RC STRAIN=C3H;
 RA Bissinger D.P., Serrero G.;
 RT Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
 CC - FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
 CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
 CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein

CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sb.ch/announce>
 CC or send an email to license@isb-sb.ch).
 CC -----
 DR EMBL; X79987; CAA5587.1; -
 DR EMBL; U03425; AAA1789.1; -
 DR EMBL; X59698; CAA42219.1; -
 DR EMBL; L06864; AAA53029.1; -
 DR EMBL; Z12608; CAA78249.1; -
 DR HSSP; P11362; 1FGX.
 DR MED; MGI:95294; Egfr.
 DR InterPro; IPR000494; EGRF_L_domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF000659; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PP01030; Recep_L_kinase; 2.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 3.
 DR SMART; SM00219; TYKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR TRANSMEMBRANE; Glycoprotein; Receptor; Signal; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
 FT SIGNAL 1 24
 FT CHAIN 25 1210
 FT DOMAIN 25 647
 FT TRANSMEM 648 670
 FT DOMAIN 671 1210
 FT REPEAT 75 300
 FT REPEAT 390 600
 FT DOMAIN 1028 1071
 FT NP_BIND 714 981
 FT BINDING 720 728
 FT ACT_SITE 839 839
 FT DISULFID 190 199
 FT DISULFID 194 207
 FT DISULFID 215 223
 FT DISULFID 219 231
 FT DISULFID 232 240
 FT DISULFID 236 248
 FT DISULFID 251 260
 FT DISULFID 264 291
 FT DISULFID 295 307
 FT DISULFID 311 329
 FT DISULFID 329 333
 FT DISULFID 506 515
 FT DISULFID 510 523
 FT DISULFID 526 535
 FT DISULFID 539 555
 FT DISULFID 558 571
 FT DISULFID 562 579
 FT DISULFID 582 591
 FT DISULFID 595 617
 FT DISULFID 620 628
 FT DISULFID 624 636
 FT MOD_RES 660 680
 FT MOD_RES 1092 1092
 FT MOD_RES 1092 1092
 CC PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 CC PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

QY	11	LLALLLPFGAA--STOVCTGIDMKRLRLPASEETHLDMRLHYOCQYVQGNLELTYLPTN	68
Db	14	LLTALCAGAGALEEKVKCGGTSTNRLTQGTFFEDHFLSLORMYNNECVLNLNLTYYQGN	73
QY	69	ASLSLFDIODEVQSVYLLIHNQVRQVPLQRLRIYRGQLPEDNATAVLNDSPLNNTTP	128
Db	74	YDLSFLKTIQEVAGVLLIHLNLTVERILPLENIQIILRGNLNENTALTALISN-----	124
QY	129	VTASPGGRELQRLSTETLEIKGVLIGRNPOLCYODITLMKDI---FKHNNQLALTLI	184
Db	125	-YGTNRTGRELPMRLQELIGAVRFSPNNPLCNMDITIGARDIVQVFMNSMSML---	180
QY	185	DTNRSAACHPCSMCKGSGRCWSESESDOSLTRYVACAGCA-RKGSLLPTDCHCECAAG	243
Db	181	-QHPSCCPKCDSPSGCSGCGGEGENQCKLTKIKCAQCSHRRGSPDCCNNQCAAG	239
QY	244	CTGPKASDCLACLHNHSGICEIHLCPALVYNTDTFESMPNBSGRYFGASCYACPNY	303
Db	240	CTGPRESDDLVCQKFODEATCKDCTPPMLLYNPPTYQMDVNPBEGKYSFGATCYKCKPRNY	299
QY	304	LSPTDVSCTLVCPLHNQEVTAEDGQRCREKSKCARCVGLGMQYIKASKRTIGTELE	363
Db	300	VVTDHSCYVRACPDYIEV-EEDGIRKRRKDDGPRKVNQIGIGERK-DTLSINATNKK	357
QY	364	-FAGCKKIFGSLAFLEPSEFDGPASNTAPLOPEQLQVETLEETGYLYISAMPDLPDL	422
Db	358	HFYKCTAIGSDHLILPVAKGDSFTRPTLPBRELILIKTYKEITGFLLIQAMPDHWTL	417
QY	423	SVKONLOVIRGRILHNHGAVALTQGLGSMWLSISLSELSGLALIHNNHLCQHNVTWM	482
Db	418	HAEENLEIIRGRKQHGQPSLAVGLNINTSGLSLKEIISGDVYISGRNNLCYANTINM	477
QY	483	DQLFRRPHQALLHTANRPDECEVGEGLACHOLCARGHMGWBPQTQVNCSSQPLRGCEVE	542
Db	478	KKLEFGPNQKTKIANNRAEKDKAVNHNCPNLCSEBGCGEPEDCVSCONVSRGEEVE	537
QY	543	EGLVLOGLPREVYNARHCLPCHPECOPOPNASVTCGPEADQVCAAHYXDPREPCVARGCS	602
Db	538	KNIIDGEREPREPNSEECIQCHPECLPDAMNITTCGRPDNDCIAHYIDGPRCVTCEA	597
QY	603	GKXPDISYMPIMKFPDEBGAQPCPPINCTHSCVDLDKGCPEAQRASPLTISYSAVGLI	662
Db	598	GIMGENNTL-VKMYKADANNVCHLCANCTYGCAGPGLOGCEVWPSGSKIPISTIGTVGGL	656
QY	663	L---VVVLGVFEGILLOYI---KANSKFIGITEU--PLTSPGAMPNOAKRIKETEELRK	714
Db	657	LFIIVVALIGGLFMRNRRIYRKRTIRLLQERELVEPLTPBGEAPNOAHRIILKETEFPK	716
QY	715	VKVLGSGAGTGYTKGIWPDGENTVKIPALIKUREBNTSPKANKEIDEAYVMAGVSPYV	774

Db	Qy	717	IKVLSSGAGTGYKGLMIPEGEKVKIPIVAILKELREATSPKANEILDEAYVVAASVDNPHV	776
Db	Qy	775	SRLLGICLTSTVQLVTQLMPFGCLLDVYRENRGLSGODLLMCMQIAGKSGTLEVRV	834
Db	Qy	777	CRLLGICLTSTVQLVTQLMPFGCLLDVYRENRKONISGYLLNVCQIAGKMLNLERRV	836
Qy	835	HRDLAARNLVKSPHHVKITDFGLARLILDDIETEHADGKVPDKMALESILRRRTHQ	894	
Db	837	HRDLAARNLVKTPGHVKITDFGLAKLLGAEXEYHAEGGKVPDKMALESILHRVYTHQ	896	
Qy	895	SDVWSYGVYVWELMTFFGKAPYDGIIPAREIFDLEKGRLEPQPICTIDVYMTWKCMMD	954	
Db	897	SDVWSYGVYVWELMTFFGKAPYDGIIPASDISILEKGRLEPQPICTIDVYMTWKCMMD	956	
Qy	955	SECPREFELVSEEFRRMARDPQRFVVIQ-NEDLGASPALDSTFYSRLLEDODGLVDAE	101	
Db	957	ADSRPFREFELVSEEFRRMARDPQRFVVIQGDGRMHLPSPTDSNFYRALMDEEDHEDVDAD	101	
Qy	1014	EYLVPPQGGFFCCDDPAPAGAGVWHRHSSSTRSGGDLTIGLEPSSEAPRSPGLASBEA	107	
Db	1017	EYLVPPQGGFF-----NSPST-----SRTPLLSSLSA	104	
Qy	1074	GSDVPEGDILGMAAKGLDLPETHDPSPIQRYSEDPTVLPSET--DGVAAPLTCSPOPEY	113	
Db	1043	TSN---NSTVACINRNGSCVKEADAFLORYSSDPTGAVTEDNIDDAFL-----PVPEY	109	
Qy	1132	VNQPDPVRPPSPRRGCPRLAPRAPGATILERAKTISPGKNQVAVDVAFGAVENPEYL-T	119	
Db	1093	VNO-SVPRKPAGSVNPFVHNQPLHP-----APGRDLHYCN--PHSNAGNPFYLLT	114	
Qy	1191	PQGAAPQCPHPAPSPAFDNLVYWDQ-----DP-----PERGAPSTFKGPT	123	
Db	1142	AQ-----PTCLSSGFNSPALWIKGSHQMSLDNPDYQDFFPKETKPNKIDFXG-DT	119	
Qy	1235	AENPEYLGSDVP	1246	
Db	1192	AENAEYLRVAPP	1203	

RESULT 6

ERBA_HUMAN

ID	ERBA_HUMAN	STANDARD	PRT	1308	AA.
AC	015303				
DT	15-DEC-1998	(Rel. 37, Created)			
DT	15-DEC-1998	(Rel. 37, Last sequence update)			
DT	15-JUN-2002	(Rel. 41, Last annotation update)			
DE	Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)				
DE	(p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).				
DN	ERBB4 OR HER4				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RN	SEQUENCE FROM N.A. (ISOFORM JM-A).				
RP	TISSUE=Breast carcinoma;				
RC	MEDLINE=93189574; PubMed=8983326;				
RA	Plozman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,				
RA	Foy L., Neubauer M.G., Shoyab M.;				
RT	"ligand-specific activation of HER4/p180erbB4, a fourth member of the				
RT	epidermal growth factor receptor family,";				
RL	Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).				
RN	[2]				
RP	SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).				
RC	TISSUE=Petal brain;				
RA	MEDLINE=97476287; PubMed=9334263;				
RA	Eleutius K., Corfas G., Paul S., Choi C.J., Rio C., Plozman G.D.,				
RA	Klagsbrun M.;				
RT	"a novel junctional membrane domain isoform of HER4/ErbB4. Isoform-specific				
RT	tissue distribution and differential processing in response to				
RT	phorbol ester.";				
RL	J. Biol. Chem. 272:26761-26766(1997).				

FT	DISULFID	308	323	BY SIMILARITY.
FT	DISULFID	326	330	BY SIMILARITY.
FT	DISULFID	503	512	BY SIMILARITY.
FT	DISULFID	507	520	BY SIMILARITY.
FT	DISULFID	523	532	BY SIMILARITY.
FT	DISULFID	536	552	BY SIMILARITY.
FT	DISULFID	555	569	BY SIMILARITY.
FT	DISULFID	559	577	BY SIMILARITY.
FT	DISULFID	580	589	BY SIMILARITY.
FT	DISULFID	593	614	BY SIMILARITY.
FT	DISULFID	617	625	BY SIMILARITY.
FT	DISULFID	621	633	BY SIMILARITY.
FT	MOD_RES	1162	1162	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1188	1188	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1258	1258	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1284	1284	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	138	138	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	174	174	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	181	181	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	253	253	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	358	358	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	410	410	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	473	473	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	495	495	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	548	548	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	576	576	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	620	620	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	626	648	NGPSSHCIYYPWGHSTLPOHA -> IGSSIEDICGLMD (IN ISOFORM UN-B).
SQ	SEQUENCE	1308 AA;	146807 MW;	5EA4809485D86761 CRC64;

	Query Match	43.0%	Score 2915.5	DB 1	Length 1308
	Best Local Similarity	44.7%	Prod. No. 4.6e-149		
	Matches 603	Conservative 183	Mismatches 377	Indels 185	Gaps 30
QY	9	WGLLLALLPGGAA----	STQVCTGTDMKRLRLPASPEHLLDMLRHLRYGCGCVOGNETLTY	64	
DB	8	WWWLTLVLAAGTV	PSDPSQVCACTEKKSSLDLEQYVHALRKYRNCCEVWGNLEITS	67	
QY	65	LPTNASLSPQIDIOE	VGYYLIAHNVQVPLQRLRVRGTOLEFEDNYALAVDNGDPLN	124	
DB	68	IENHRDLSFRSAREV	LYVALNORFVPLENRLIRGTXIYEDRYALAFILNPKDG	127	
QY	125	NTTPVTSAGPGSLRE	LOLRSLTEILKGGVLIQNNPOLCYODTILMKQIPIKQNNLATLI	184	
DB	128	NF-----	GIQELGKRLTELLNGVYVDQNKFLCYADTIHMQIVANPMPNSLTLV	178	
QY	185	DTNRSRACHPGSPNCK	SGSRCEWSSSEDCSLRTVACGC-APCKGPLPTDCHBOCAAG	243	
DB	179	STNGSSGCGRCHKS	CTG-RCMGPTEHNCQTLLRTVCAEQDCDGCYGGVYVDDCCHRECAAG	237	
QY	244	CTGKRHSDCLAC	LHPNNSGICELHCPALVLYNDDTFPSMNPBGRVYFGASCYTACPNY	303	
DB	238	CSGKRDIDCAKNN	FNDSGACVYQCQPTPYNNPTTQLENNPAKTYTGAFVYKCKPHNF	297	
QY	304	LSTDVSGCTLVCP	LHNOEVTAEADGTORCEKSKPCARVCYGLQMYYIKANSKEFIGITELE	363	
DB	298	V-VDSSCVYRAC	SSRMVEV-EENGIMKCKPCTDICPKACGIGITGSLMSKQYDSSINIDK	355	
QY	364	PACCKKIFGSLA	LPESFGDPSANTAPLOPEOLOVPELTETIGLYLISAMPDGLPDIS	423	
DB	356	FIMCTKINGNLI	LVLVYGHDPYNALBALDPEKLVNFRVYRELTIGLPLNQSMPNNTDS	415	
QY	424	VFQNLQVIRGRIL	HNQAVSLTLQGLGISWGLSRLELSGALLHHNTHLCFVHTVPWD	483	
DB	416	VFNVLVTIGR	VLVYSGJSLILKQOQITISLOFQSLKEISAGNIYITDNSNLCYVHTINMT	475	
QY	484	QLFRNPQALHTP	ANRPEDECVAGEGLACHQLARGHGMWGGPQCNAGCOPLRGQSCVEE	543	
DB	476	TUJSTINQRI	VIRIDNKAKENCTRAEGVACHLSSDQGWGPGPQCLSCRSFRSGRLCIS	535	
QY	544	CRYLQGLPREY	VARHCLPCHPECPQ-QNGSVYTCFQPEADQCVAAHYKDPFCVYARCS	602	


```

FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 CYS-RICH.
FT DOMAIN 718 985 PROTEIN KINASE.
FT NP_BIND 724 732 ATP (BY SIMILARITY).
FT BINDING 751 751 ATP (BY SIMILARITY).
FT ACT_SITE 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 PYT -> SYR (IN REF. 3).
SQ SEQUENCE 1308 AA; 146957 MW; D944BB0996A08B41 CRC64;

```

Query Match 42.7%; Score 2896; DB 1; Length 1308;
 Best Local Similarity 44.6%; Freq. No. 5,1e-148;
 Matches 601; Conservative 190; Mismatches 386; Indels 172; Gaps 30;

```

QY 1 MEIA-ALCRWGLL--ALLPFGAASVCTGTDMKLRLPASPEETHLMLRLHYGCGGVQ 57
Db 1 MKLATGLWVGSLVAARTVOPSASQSCAGTENKLSLSLEQYALRKTYENCEVM 60
QY 58 GULELTYLPTNASLFLDIOGVGVLIANNQVQVLRQLRVGTQLEEDNYALAVL 117
Db 61 GULELTSIEHNRLDFLRSIRVETGVVALNPFVLEMLRIIRGKLYEDRYALAF 120
QY 118 DNGDELNNTTPTVTCASPGSLRELOCSLTIELIKGVLIQBNPQLCYDITLMKDIHKFN 177
Db 121 LNYRKDGNF-----GLQELGAKNTEILNGSVYDQKFLCYADVTTHWQDIYANPW 171
QY 178 QALATLIDTNRSRACHPCSPMKCGSRGSESSDQCSTLRITVCAGGC-ARCKGFLPTDC 236
Db 172 PSNMTLVSTIGSSGGRCHKSCGTG-RCWGPTEHHCQTLTRITVCAEQCDGRCGYVSDCC 230
QY 237 HECCAAAGCTGPRHSCCLALPHNHSIGELACPLALVYNTDTFESMNPBGRITTFGASCV 296
Db 231 HRECAAGCSGRKDTICFACNMNPNOSGACVTCQPOTFYNPPTFLFNHNAKXTYAGFCV 290
QY 297 TACPNYVLTSTVSGCTLVCPLHNOEVTAEQTORCEKSKFCARVAVCYGLMQYIKANSKF 356
Db 291 KKCPHNFV-VDSSSCVCRACPSKMEV--EENGITKMKCRCTDICPAACGIGITGSLMAQIV 348

```

```

QY 357 IGTELEFAGCKKIFGSLAFLEPESFDGDPASNTARLOPEQLQVETLEETIGYLYISAWP 416
Db 349 DSNNTKFNCTKINGNLIFLTGTHGDPYNAIDAIDEKLVNFTVEITGFLNIQTW 408
QY 417 DSLPLSVFQNLQVIRGRILHNGAVSLTLOGISLWGLRSRELGSALAHNTHLCF 476
Db 409 PNMTRDSVSNLVTIGRVLVSGLSLLIKQGGISLSPQSLKEISACNIIYITDSNICY 468
QY 477 VHVVPDQLEFRPHQALHTARPEDECVGSEGLAQHQAAGHCGPPTQCVCNCSQFLR 536
Db 469 YHTINMTLTFSTVNRIVIRDRRAENCTABGMVGNHLCSDQGCWGPDPDCLSCRRFSR 528
QY 537 GQECVEECVQLQGLPREYVNAHRCPCPCPECP-ONGSVTCFCEPADQCVACAYKDPF 595
Db 529 GKICIESCNLYGGEFRREFNGSICVECCSQCKEMEDGLTCHGPRDNTCASHFKQPN 588
QY 596 CVARCDGVKPLSTWPIWKPDEEGACQPCPNCTHSCVDLDDKC-----PA 644
Db 589 CVEKCPDVLQGANSF--IFKYADQRECHPCHPNCTQCNGPTSHDCIYYPWGHSTLPQ 646
QY 645 EGRASPLSVSAVY-GILLVVLGVGVGILLIYIKANS-----KFIGITEL--PLT 693
Db 647 HAR-TPL--IAAGVIGSLVITVMTALTAV--YVRRSIKKKALRRFLE-TLVEPLT 699
QY 694 PSGAMPNQAOMRILKETEELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVRENTSP 753
Db 700 PSQTAAPNQALMLKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVRENTGP 759
QY 754 KANKELDEAVYMAVGSPYVSRLLIGLTSTVQLVQLMAYGCLLDHVRNRRGLCSQD 813
Db 760 KANVERMEALIMASVDPHPLVRLGVLSPTLQVQLMPHGCLLEVHVKXNISQOL 819
QY 814 LNMGCQIAKAGSYLEDRLVHRDLAARNVYKSPNHKIDFGRLALDDETEYADG 873
Db 820 LNMVCQIAKAGMYLEERLVRHDLAARNVYKSPNHKIDFGRLALDDETEYADG 879
QY 874 GVPLIKMALESILRRFTTHOSDVMSYGVYTWELMTFGAKPYDGIIPAREIDLLEKERT 933
Db 880 GMPPIKMALECIHRRKFTTHOSDVMSYGVYTWELMTFGAKPYDGIIPAREIDLLEKERT 939
QY 934 POPPCTIDVYVIMYKCMIDSEORPRELVSFSSMARDPQRFVYIQMED-LGPASPL 992
Db 940 POPPCTIDVYVIMYKCMIDSDSPKFEKELAAFSRRARPPQRYLVIQGDDRKLPSPN 999
QY 993 DSTFRSLLEDMDLVDAAEYVLPQGFPCPD-----AP 1029
Db 1000 DSKFQNLLEDDELDMDDAEYVLP-QAFNIPPIYTSKTRIDSNSEIGHSPRAYTP 1058
QY 1059 GAGGMYHRRHSSSTRSGGDLTGLBSEEEAARSLAPSEAGSDVPFGDLAMCAKG 1089
Db 1059 MSGGFVYQDGFATQCG--MPMPYATATSTIPEAVA--QGATAEMFDDSCNGTLRK 1113
QY 1090 LQSLPTHPSPLQFRTSEDPVPLPS-----ETDGVVAPLTCSPQEEYVNOPDVRQPP 1142
Db 1114 FVYHVAVQDSSSTQKTSADPTVFAERKPRPAELDEEGIMTGMKDKPKKEYLNPNVE----- 1167
QY 1143 SPREGPLPAARPAQATLERAKTSLPGKNGVYKDFAFAGAVENPEYLLTPQGAAPQPHRP 1202
Db 1168 ---ENPFVSRR-----XNGDLC-----ALDNPETHASASG-----PP 1196
QY 1203 PA-----PSAPFDLVLVYDQPPPRGA--PP 1226
Db 1197 KADEEYVNEPLVNTFTNALGNAEYMKNSLSPREKAKKAFDNPDPWNHSLPRSTLQHP 1256
QY 1227 STFKGTPF-----AENPEYL 1241
Db 1257 DYLOEYSTKYFYKONGRIRPIVAENPEYL 1285

```

RESULT 8
 XMRK_XIPMA
 ID XMRK_XIPMA STANDARD; PRT; 1167 AA.
 AC P13388;

DT 01-JAN-1990 (Rel. 13, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
 GN XMRK OR TU.
 OS Xiphophorus maculatus (Southern platyfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorphi; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OX NCBI_TaxID=8083;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90015140; PubMed=2797166;
 RA Wittbrodt U., Adam D., Malischek B., Maueier W., Raulf F.,
 RA Telling A., Robertson S.W., Schartl M.;
 RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
 RT inducing Tu locus in Xiphophorus.";
 RL Nature 341:415-421(1998).
 RN [2]
 RP REVISION TO 515.
 RA Schartl M.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 CC - FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC - SUBCELLULAR LOCATION: Type I membrane protein.
 CC - DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X16891; CAA34770.2; -.
 DR PIR: S06142; S06142.
 DR HSSP: P1362; 1FGK.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00069; kinase_1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR PRINTS: PRO109; TYRKINASE.
 DR ProDom: PDO00001; Euk_pkinase; 1.
 DR SMART: SMO0261; FU; 5.
 DR SMART: SMO0220; S_TKC; 1.
 DR SMART: SMO0219; TYRK; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
 FT SIGNAL 1 25
 FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE
 FT KINASE.
 FT EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 643 665 POTENTIAL.
 FT DOMAIN 666 1167 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 710 977 PROTEIN_KINASE.
 FT BINDING 743 743 ATP (BY SIMILARITY).
 FT ACT_SITE 835 835 ATP (BY SIMILARITY).
 FT DISULFID 195 204 BY SIMILARITY.
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 220 228 BY SIMILARITY.
 FT DISULFID 224 236 BY SIMILARITY.

FT DISULFID 237 245 BY SIMILARITY.
 FT DISULFID 241 253 BY SIMILARITY.
 FT DISULFID 256 265 BY SIMILARITY.
 FT DISULFID 269 296 BY SIMILARITY.
 FT DISULFID 300 311 BY SIMILARITY.
 FT DISULFID 315 330 BY SIMILARITY.
 FT DISULFID 333 337 BY SIMILARITY.
 FT DISULFID 504 513 BY SIMILARITY.
 FT DISULFID 508 521 BY SIMILARITY.
 FT DISULFID 524 533 BY SIMILARITY.
 FT DISULFID 537 553 BY SIMILARITY.
 FT DISULFID 556 569 BY SIMILARITY.
 FT DISULFID 560 577 BY SIMILARITY.
 FT DISULFID 593 615 BY SIMILARITY.
 FT DISULFID 618 626 BY SIMILARITY.
 FT DISULFID 622 634 BY SIMILARITY.
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1167 AA; 129934 MW; 4799E0749DC1D55A CRC64;
 Query Match 38.6%; Score 2612.5; DB 1; Length 1167;
 Best Local Similarity 44.3%; Pred. No. 7,2e-133;
 Matches 561; Conservative 167; Mismatches 394; Indels 145; Gaps 30;
 QY 4 AALCRWGLLALLPFGAAS-----OVCGTGTMKRLPASPERHLDMLRLHVGQGVQVGN 59
 DB 8 AALLQ--LLVLVLSIRCCSTDPDRVCOGTSGNQTM--LNNHLYMKMKMISGCVNLEN 62
 QY 60 LLITLPTNASTLSPLQDIQEVGYVLAHNOVROVPLQRLRIVEGTQYFEDNVALVLDN 119
 DB 63 LLITLTOENODLSFLQSIQEVGYVLAHNEVSTPLVNLRLIRGQNLVEGNFLVWMSN 122
 QY 120 GDPLNNTPTVPGASPGSLRELQSLSTFELKGVLLIQRPCLCYQDTILMKDIFHKNNQL 179
 DB 123 YK--NPSSP--DIVYGVKQJQLQSLVLEISGVKVSINPLLCVETINMWDDVDKTSNP 179
 QY 180 ALTLIDNRSRACHPCSPMKSGKSGWGSSEDCSLRTVCAAGC--ARKSGPLTPDCHE 238
 DB 180 TNNLIPHAFERQCKCKCHGCVNGSOMAPGPHQCKFKLLCAEQNRRCRGPRIIDCNE 239
 QY 229 QCAAGCTGPKASDCLACHFNHSGICELHCPALVTYNTDFESNPNEGRTYFGASCVTA 298
 DB 240 HCAGGCTGPRATDCLACRFDDDCDCTGCPPKIYDIVSHQVVDNENIKYTFGAACVKE 299
 QY 299 CPVNLSTDVSGCTLVCPILHNOETAEADGTORCKSGKPCARVGVYGMGYIANSKFIG 358
 DB 300 CPNSNVVTE--GACVRSAGMTEVD--ENGKRSCKPCDGVCPKCDGIGISL--SNTIAVN 356
 QY 359 ITEL--EPGCKKIKFGSLAFIPESFDGDPASNTAPLQPEQVETLEELITGYLYISAMP 417
 DB 357 STNIRSFENCTKINGDIIILNRNSPEEGPHYKIGMDPEHLMNLTVEITGVLVIMWPE 416
 QY 418 SLPLSVFQNLQVRRGLINNGAAS--LTQGLGISTWGLSLRLSLGSLALHNNTHLCE 476
 DB 417 NMTLSLSPQNLLEIRGTTFSRGSSFVAVVRHLQWGLSLSLKEVAGNVLNNTQLRY 476
 QY 477 VHTVPMDQLFENPQALHTANRPEDECVEGGLACHQLCARGCHWPGPFCVNGSQPLR 536
 DB 477 ANTIMRRRLFRSEDSQISEYDART-----ENGTNNCEGSDGCMGPGPMCVSCLHVR 529
 QY 537 GQEVVEECRVLQGLIPREYVNAKCLPHEPCQFQNSVTCFGEADQCVAAAHYKDPFC 596
 DB 530 GGRCVASCNLLQEPREAOVDGRCVQCHOECLVQDTSLTCYGGPANGSKSAHFQGGPQC 589
 QY 597 VARGSPGVKPDLSYMPIMKFPDEGACQPCPINCSTCVDDLDKGPAGQASPLTISVS 656


```

FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.
FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 286 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 552 565 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VASPPIC 141 183 EILSGVYIEKNDKCHMDTIDMDRDIIVDRDAEIVKONGR
SC -> GGFPMFSGSLTPOADWYLLDDDDPRLLLSASK
VPTLAAY (IN SHORT ISOFORM).
FT VASPPIC 184 1342 MISSING (IN SHORT ISOFORM).
FT CONFLICT 560 560 E -> G (IN REF. 2).
FT CONFLICT 1064 1064 E -> G (IN REF. 2).
SQ SEQUENCE 1342 AA: 148097 MW: 7201E7F6CA374ED CRC64;

```

Query Match 35.3%; Score 2390.5; DB 1; Length 1342;
 Best Local Similarity 40.2%; Pred. No. 6.8e-121;
 Matches 527; Conservative 197; Mismatches 449; Indels 139; Gaps 35;

```

QY 10 GILLALLPFGAA--STQVCTGTDMKRLRASPETHLDMKRLHYOCGVQGNLEITYPT 67
DB 11 GLPFLARSGSEVGNISQAVCPGLTNGLSVTDANQYQTLKYERCEVMKNLEIVLGH 70
QY 68 NASHSLQDIQEVQVYLIAHQVROVPLQRLIYRGTLFEDNALAVLDNGPLNNTT 127
DB 71 NADSLFLQWIRVLTGYLVAMNEFTLPLNLRVVGQVYDGKFAIVM-----LNYNT 125
QY 128 PVTGASPGGLREIQLSLTILKGVLIQNPOLCYQDTILMKDIFKQNLATLTDN 187
DB 126 ---NSSHALRQLRLTQLTEILSGVYIEKNDKCHMDTIDMDRDIIVDRD--AEIVKD 178
QY 188 RSRACHPCSPMKSGSRWGSSEDCSLTRTVACAGC-ARCKGPLPTCCHEOCAGCTG 246
DB 179 NGRSCPCPCHEVCKG-RCKGPGSEDCQTLTKTICAPQCNHCGGPRNQCCHDECGSG 237
QY 247 PKISDCLACHENHSGICELHCPALTYMTDFESHPNBEGYTFGASCYTAAPRYNT 306
DB 238 PDDTDCFACHFNDISACVPRCPQPLVYNKLTFFOLEPNHTTYQVGGCVASCPHNFV-V 296
QY 307 DVSGSLVCELANQVETABDGTQRCCKSPCARVYGLMGQYIKANSKF--IGITEL- 363
DB 297 DQTSVACAPPDMEVD-KNGLMKCEPCGGLCPKACEGSG-----SGSRFQTVSSNIDG 350
QY 364 FAGCKRTFGSLAFLPESFODDPAISNAPLQPELOVFELEITTYLYTSMWPSLPDLS 423
DB 351 FVNCCTKLGNLDFLTGLNGDPMHKLPAIDPEKLNFRVREITTYLYTSMWPSLPDLS 410
QY 424 VFQNLQVIRGRILHNGAVS-ITLQGLISMLGRLSRLSGLALIHNTLCEVHTYPM 482
DB 411 VFSNLTITIGRSLYNGFSLIMKLNLTSLGFRSLKESNRIRITYISANROLCTVHSLNM 470

```

```

QY 483 DQLFPHQALLHTA-NRDECEVGEIACHOLCARGHCWGPFGTQCVNCSQFURGQECV 541
DB 471 TKVLGRPEERLDIKHNRPRRDCVABEKVCDPLSSGCGCPGPGGCLCRNRSRGVY 530
QY 542 EECRVLOGLPEEYVNAHCLPCHECOPONGSVTCTGFPEADQCVACAHKDPFCYARCP 601
DB 531 THCNELNPEPEFNAECSCHEPCQPMGTATCNGSGSDTCAQCAHFRDGHCVSSCP 590
QY 602 SGVXKDLVYMIWMPPEBEGACGCPINTCHSCVDJDDKCAPAEORA-----SPLTSIVA 657
DB 591 HGVLG--AKGFIYKTPDYNECRCHENCCTGCKGPELDQCLQTLVLGKTLTALTV 648
QY 658 VVGLLV-VLVGVF---GILQYIKANSKF--GITTEPLTPSGAMPNQAQMLKET 710
DB 649 IAGLVFIMMIGTGFLYWRGRILCNKRAMRYLERGESIEPLDPS--EKANKVLARIKFE 707
QY 711 ELRKXVLGSGAFGVYKGIWIPGENVVKI PVAIKYLRRENTSPKAKELIDEAYVAVAG 770
DB 708 ELRKXVLGSGAFGVYKGIWIPGENVVKI PVAIKYLRRENTSPKAKELIDEAYVAVAG 767
QY 771 SPYVRLIGLITSTVQLVQMLPFGCLDHYENRGRIGSQDLNMCQIAKMSYLED 830
DB 768 RAHIVRLTGLCPGSSSLQVLVQYPLGSLDHYVQRGALGPQLLNMVQVIAKMYLLE 827
QY 831 VLVYRDLAANVLYKSPNHYKIDPGLARLDIDETRYADGKVPKIMMALESILRR 890
DB 828 HGVYRDLAANVLYKSPNHYKIDPGLARLDIDETRYADGKVPKIMMALESILRR 887
QY 891 PTHQSDVSYGVYTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDYVMIVK 950
DB 888 YTHQSDVSYGVYTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDYVMIVK 947
QY 951 WMLDSECPRELVSESRRAARPOCFVYIYQNEIDGFA---SPLDSTFRSLLEDDMG 1007
DB 948 WMLDSECPRELVSESRRAARPOCFVYIYQNEIDGFA---SPLDSTFRSLLEDDMG 1006
QY 1008 DLVDAEVLVPOQGFPCDPAPAGAGVYHRRHSSSTRSGGDLTGLP--SEEAAPRS 1066
DB 1007 PELDLDLLEED-----NLATTTIGSLSLPVGTLNRPQSGSL 1047
QY 1067 LAPSGASDVFDGDLGMAKGLQSLPTD--PSPLQRYSEDTVLP-----SETDGY 1119
DB 1048 LSPSSGY-MPNQNLGSESCQESAVSSSERCPRPSLH-----BMPGCLASESSEGH 1100
QY 1120 VA-----PLTCSQPE-----YNAQPVRRPQPSPEGP----- 1148
DB 1101 VTGSAELVQEKVSMCRSRSSRSRRPGRDSAHYSQKHSILTPYTPSPGLSEEDVNGY 1160
QY 1149 LPAAPAGATIERAKTISP-GKNGVY-----KDYAFGAVENPEYLPQGCAPQHP 1201
DB 1161 MDTLAKGTSPSRGTLSSVGLSGLTEBEDD-----EEVEYMRRRRHSF-PHP 1211
QY 1202 PPAFPAPFNLVYWD-----QDPEFGAPSTRTKGTPTAENBYL 1241
DB 1212 PRPSLEELGVEYWDVSDLSASLSTQSCPLRPVIMRAGTTPBEDYEM 1263

```

RESULT 10
 ERB3 RAT
 AC 062759; 062955;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
 GN (c-erbB3).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eumetazoa; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RX MEDLINE=96096535; PubMed=8522190;
 RA Heiliger N.J., Kim H.-H., Greaves C.H., Sterke S.L., Roland J.G.;
 RT "Cloning of the rat ErbB3 cDNA and characterization of the
 RT recombinant protein.";
 RL Gene 165:279-284 (1995).
 RN
 RP REVISIONS TO 85, 513 AND 565.
 RA Heiliger N.J., Roland J.G.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RN
 RP SEQUENCE OF 922-1097 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
 RX MEDLINE=97184212; PubMed=9030624;
 RA Carroll S.L., Miller M.L., Fromert P.W., Kim S.S., Corbett J.A.;
 RT Expression of neurofilins and their putative receptors, ErbB2 and
 RT ErbB3, is induced during Wallerian degeneration.";
 RL J. Neurosci. 17:1642-1659 (1997).
 CC
 CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTK.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
 CC PHOSPHATIDYLINOSITOL 3-KINASE.
 CC
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb.sib.ch).
 CC
 CC EMBL: U29339; AAC8498.2; -
 DR EMBL: U52530; AAC53050.1; -
 DR HSP: P11362; 1FGK.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR PRINTS: PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; FU; 5-
 DR SMART: SM00219; TYRC; 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; FALSE_NEG.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 19
 FT CHAIN 20 1339 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
 FT DOMAIN 20 643 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 644 662 POTENTIAL.
 FT DOMAIN 663 1339 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 183 259 CYS-RICH.
 FT DOMAIN 707 964 PROTEIN KINASE.
 FT NP_BIND 713 721 ATP (BY SIMILARITY).
 FT BINDING 740 740 ATP (BY SIMILARITY).
 FT ACT_SITE 832 832 BY SIMILARITY.
 FT DISULFID 186 194 BY SIMILARITY.
 FT DISULFID 190 202 BY SIMILARITY.
 FT DISULFID 210 218 BY SIMILARITY.
 FT DISULFID 214 226 BY SIMILARITY.
 FT DISULFID 227 235 BY SIMILARITY.

FT DISULFID 231 243 BY SIMILARITY.
 FT DISULFID 246 255 BY SIMILARITY.
 FT DISULFID 259 286 BY SIMILARITY.
 FT DISULFID 290 301 BY SIMILARITY.
 FT DISULFID 305 320 BY SIMILARITY.
 FT DISULFID 323 327 BY SIMILARITY.
 FT DISULFID 500 509 BY SIMILARITY.
 FT DISULFID 504 517 BY SIMILARITY.
 FT DISULFID 520 529 BY SIMILARITY.
 FT DISULFID 533 549 BY SIMILARITY.
 FT DISULFID 556 573 BY SIMILARITY.
 FT DISULFID 576 585 BY SIMILARITY.
 FT DISULFID 589 610 BY SIMILARITY.
 FT DISULFID 613 621 BY SIMILARITY.
 FT DISULFID 617 629 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 1028 1028 L -> P (IN REF. 3).
 SQ SEQUENCE 1339 AA; 147545 MW; 0AA5FF2402B8DF1E CRC64;
 Query Match 34.2% Score 2316.5; DB 1; Length 1339;
 Best local similarity 40.5%; Pred. No. 63e-117;
 Matches 520; Conservative 173; Mismatches 427; Indels 163; Gaps 36;
 QY 3 IALCRWGLTALLPPGAA---STQVCTGTDMLRLPASPETHLDMRLHYOGCQVVOGN 59
 DB 7 LQVLC-----FLSLIARGSEMNSQVCGTLNGLSVGDADNQYQLYKYEKEVEVMGN 62
 QY 60 LELTYLPNNAISPLQDIQEVGYVLIANQVQVPLQRLRYNGTQFEDNYALVLDN 119
 DB 63 LEIVLTGNADISFLQWIREVTGYLVANMEFSVLPFNLRVARGTVYQGGKFAIFWM-- 120
 QY 120 GDLPLNNTPTVYGASPGGLREQLSLTEILKGVLIQRNPOLCYQDTILKMDIFHKNNQL 179
 DB 121 ---LWYNT-----NSSHALRQLKFLQTEILISGVYIERKDKLCHMDIIMRDIVRA--- 170
 QY 180 ALLTIDNRSRACHPCSPCKGSGKSGWESSDCSLRTVYAGGC-ARCGKPLPTDCHE 228
 DB 171 GAEIVVKNNGANCFPCHEVCKG-KWCGGPPDDQILTKTICAPQCNQRCFGPNMQCCHD 229
 QY 239 QCAAGCTGPKHSDCLACHFNHSGICELHCAVLTNTDPEESNPNEGRTGASGVTA 228
 DB 230 EAGAGCGSPQDTDFACRRFNDGACVPRCEPLVYNNKLTFLQLEPNPHTRYQYQGVCVAS 289
 QY 239 CFVNYLSTDVSGCTLVCPRLHNOETABDGTORCEKSGKPCARVYCYL--GMQYIKANSKF 356
 DB 230 CFHNHV-VDQTFECVACRPPDMEVD-KHGLKMCBPCGGLCPKACEGIGSSSRVQYVDSN 347
 QY 357 IGIITLEAFAGCKKIFGSLAFIPESFDDPASNTAPLPQLOVPELTLEITGYLYISAMP 416
 DB 348 ID---GVNCTKILGNIDFLITGLANDPMHKIPALDPEKLANFRYRELTGYLNIQSWP 403
 QY 417 DSLPLSYFQMLQVIRGLIHNGAYS-LTQGLGISWLGSRSLRELSGLALIHNTHTLC 475
 DB 404 PHMNFVSFSLUTLTIIGRSLYNRFSLIKMKNLVTSLSGRSLKEISAGVYISANQQLC 463
 QY 476 FVHYTPMQLFRNHQALLHTA-NRPDECVGSLACHOLCARGHCMGPGPTQVNCVSGF 534
 DB 464 YHSHLMTWRLRLRGSBERLDIKYDRPLGECIAGSKVCDPLCSSGGCGPQGLCLCRNY 523
 QY 535 LRQGEVCEKCVLQGLPREVYNABHCLPGRECPQNGSVTCGPREADQVAAHYAKDP 594
 DB 524 SRBGVCTHCHFLQGEPRFVHEHQCSCHPECLPMEGISTCGSGSDAARCAHFRDGE 583
 QY 595 FCVARCBGQVPRDLSTYMPIMKFPDEBGAQCPCEINCTHSC-VLDLDDKGCAPQARAPL 652

```

Db      584 RCVNSCPHSTLGA- AKGPIYKYPDQNECRPCHENCTGCGNPELQDLCGAELMSKPH 641
Qy      653 STSAVAVGI-LLVVLGVVF-----GILIQYIKANSKFI--GITEPLTPSGAMPNQAMR 705
Db      642 LVIAVTAGLAVILMLIGSGFLVWRERRIQNKRAKRYIERESIEPLDPS- EKANKTLAR 700
Qy      706 ILKETELRKVKVLGSGAGCTGYKGINIDGENVKI PVAIKYLRNTPSKANKETLDEAVY 765
Db      701 IFKETELRKVLKGGVGTGVAHKGIWIEGESIKI PVICIKYIEKSGRSOSFOATDMLA 760
Qy      766 MAGVSPYVSRLLGICLTSTVQLVQLMPYGCCLDHYVENGRGLGSDLLMCMQIKAM 825
Db      761 VCSLDHAHIVRLGLGCPSSQLDTVYQLPGLSLDHKQKHETIGPQLMLNMGVQIKAM 820
Qy      826 STLEVRVLRDPLAARNVLVKS PNHVKITDPLGLRLDIDETEHADGKVPKIMWALS 885
Db      821 YLLEHSHVHRDLARNLMLKSPSQVQVADPGVADLLPDDKQLHLSEAKTPIKMWALS 880
Qy      886 ILRRFTHOSDWSYGVTVWELMFGAKPYDGIAPAREIPDLIEKGERLLPQPPICTIDVYK 945
Db      881 IHFGKTHOSDWSYGVTVWELMFGAPYAGLRLAETPDLIEKGERLAPQICTIDVYK 940
Qy      946 IMVKCMWIDSECRPFRELVSFSSMADPQRFVVIQNEIDYGPASPIDSTFYRSLIEDD 1005
Db      941 VVVKCMWIDENIRPFEKLANEFEMADPPRYLVIKAS- GPETP---PAPESVLTKE 997
Qy      1006 MGDVDAEEVLYPQCGFFCPDPAGAGMWHHRSSSTSGGSDLTLGLEPSE 1060
Db      998 L-----QEALELEPLD-----DLDLDEAEEGSLATS 1023
Qy      1061 -----EAFSPSLAPSEG-----AGSDVFDGLGWAAGKQSLPTHD 1097
Db      1024 LGSALSLPTGTTLTPRSGQSLSPSSGVMNNQSSLGECALDSVLAGRGCFRPLSLH- 1082
Qy      1098 PSLPLQRYSEDPTVPLPSTDTGVY---APL-----TC-----SPQPE---YVYQPDV 1137
Db      1083 PIPGR-----PASESSEGHVTGSEAELOEKVYCSRSRSRSPRPDSDAYHSQRHS 1135
Qy      1138 RPOPPSPREP-----LPARPAGATLERAKTLSP-QNGVY-----KVFAF 1179
Db      1136 LITVTPISFPGLEEDONGVMPDTHLRHASSREGITLSSVGLSTGEDEED----- 1191
Qy      1180 GGAVERNPEYLTPOGGAAPQPPHP 1202
Db      1192 -----EYEVWNRKRGRSP-PRPP 1209

RESULT 11
EGRF_DROME STANDARD; PRT; 1426 AA.
ID EGRF_DROME PRT; 1426 AA.
AC P04412; O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EGF 2.7.1.112) (Egfr)
DE (Garten receptor) (troped protein) (Drosophila relative of ERBB).
GN EGRF OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephyridioidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schubach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT subdomains of the receptor protein."
RL Genetics 137:531-550(1994).
RN [2]
RP REVISIONS.

```

```

RA Clifford R., Schubach T.;
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlesinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains."
RL Cell 40:599-607(1985).
RN [4]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RX STRAIN=Oregon-R; TISSUE=embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcripts."
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RX Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RA "Several levels of EGF receptor signaling during photoreceptor
RT specification in wild-type Ellipse, and null mutant Drosophila."
RL Dev. Biol. 205:129-144(1999).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RX STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayani A., An H.-J., Andrews-Plambeck C., Baldwin D.,
RA Ballew R.M., Baau J., Bendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernon B.P., Bhattacharjee D., Bolshakov S.,
RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Butris J.M., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Chertys J.C., Cavley S., Dahle C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Maye A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Murzy D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacible J.M.,
RA Palazolo M., Plitman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Sanders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Giths R.A., Myers E.W., Rubin G.M., Venter D.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [7]
RP SEQUENCE OF 959-1078 FROM N.A.
RX STRAIN=DaekwanYeong;
RX MEDLINE=8517938; PubMed=2983232;
RA Madsen S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RA "A Drosophila genomic sequence with homology to human epidermal
RT

```



```

RT growth factor receptor."
RL Nature 314:178-180(1985).
RN [8]
RP SEQUENCE OF 1133-1137, 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
RX ANALYSIS.
RA MEDLINE=92038942; PubMed=1936959;
RA Raz E., Schejter E.D., Shilo B.Z.;
RA "interallelic complementation among DER/flb alleles: implications for
RT the mechanism of signal transduction by receptor-tyrosine kinases."
RL Genetics 129:191-201(1991).
RN [9]
RP REVIEW.
RX MEDLINE=97248481; PubMed=9094709;
RA Perlmutter N., Perkins L.A.;
RT "there must be 50 ways to rule the signal: the case of the Drosophila
RT EGF receptor."
RL Cell 89:13-16(1997).
CC - FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
CC MARK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOSEROSA
CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
CC CUTICLE.
CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC - SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
CC PROTEIN.
CC - ALTERNATIVE PRODUCTS: 3 ISOFORMS, TYPE I (SHOWN HERE), TYPE II AND
CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC - TISSUE SPECIFICITY: UNICITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
CC TESTIS AND OVARY AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FORK, LEVELS REMAIN HIGH
CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
CC AND THORACIC AND ABDOMINAL GANGLIA.
CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL AF052754 AAC08536.1; -.
DR EMBL AF052753 AAC08536.1; JOINED.
DR EMBL AF052754 AAC08535.1; -.
DR EMBL AF052752 AAC08535.1; JOINED.
DR EMBL K03054 AA551462.1; -.
DR EMBL K03417 AA551460.1; -.
DR EMBL K03416 AA550985.1; -.
DR EMBL K03418 AA551461.1; -.
DR EMBL AF109077 AAD26134.1; -.
DR EMBL AF109078 AAD26132.1; -.
DR EMBL AF109082 AAD26132.1; JOINED.
DR EMBL AF109078 AAD26133.1; -.
DR EMBL AF109084 AAD26133.1; JOINED.
DR EMBL AF109079 AAD26130.1; -.
DR EMBL AF109081 AAD26130.1; JOINED.
DR EMBL AF109079 AAD26131.1; -.
DR EMBL AF109083 AAD26131.1; JOINED.
DR EMBL AF109080 AAD26135.1; -.
DR EMBL AE003454 AAF46732.1; -.
DR EMBL X02293 CA26157.1; -.
DR EMBL X78920 CA555523.1; -.
DR EMBL X78918 CA555521.1; -.

```

```

DR EMBL X78919 CA555522.1; -.
DR PIR A00640: GOFPE.
DR HSSP P11362: IEGK.
DR Flybase: FBgn0003731; Egfr.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR002174; Tyrosin-kinase.
DR InterPro: IPR001245; Tyr_Pkinase.
DR Pfam: PF00069; Pkinase; 1.
DR Pfam: PF00757; Pkin-1-like; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_Pkinase; 1.
DR SMART: SMO0261; FU; 7.
DR SMART: SMO0219; Tyrc; 1.
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE: PS00109; PROTEIN KINASE_TYR; 1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane: Glycoprotein; Receptor; Phosphorylation; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
KW Developmental protein.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1426 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT DOMAIN 31 868 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 869 889 POTENTIAL.
FT DOMAIN 890 1426 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 938 1198 PROTEIN KINASE.
FT NP_BIND 944 952 ATP (BY SIMILARITY).
FT BINDING 971 971 ATP (BY SIMILARITY).
FT ACT_SITE 1063 1063 BY SIMILARITY.
FT MOD_RES 902 902 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).

Query Match 28.6%; Score 1935; DB 1; Length 1426;
Best Local Similarity 32.4%; Pred. No. 1.9e-96;
Matches 463; Conservative 183; Mismatches 428; Indels 354; Gaps 41;

24 QVCTGTDMKRLPASPETHDMLRLHLYQGCVQVGNLELYLP-NASLSPFDIDIOEVG 82
100 KICIGTKRSLVSPKKEHYNLRDRYNTCTYDGNLKLTLPEMNDLSFLPNIRPVTV 159
83 YVLAHNVQRPVLPRLRIYRGTLF-----EDNVALAVDNGDPLNNTPTGASPGCL 137
160 YILSHVYKRVKVPKQIIRGLTFLSLSEBEKALFV-----TISKM 203
138 RELQIRSLTEILKGVLIQRNPQCYDITLWKQIFKNNQALTLIDTRSPACHPSP 197
204 YLLEIPDLRLVNLQVGFHNNYNNCHMTIOWMSIIVNGTDAYNYDFTAPRECEPKCHE 263
198 MCKGRCKWGESSEDCQSLTRTVCAAGGA--RCKGRLPTDCHEGCAAGCTGPRHSDCLAC 255
264 SCTHG-CWGEPRKCKQKFSKLTGSPQCGAGRCYCFKREBCCHLFCAGAGCTGPPQKDCIAC 322
256 LHFNHSIGCELAHCPALVTYNTDTPESNPPEGRYTFGASCCTACPVNYLSDVGSCTLVC 315
323 KNFPFENAVSKCECPMRKYNPTTVLLENNPEGKAYATCYKECP-CHLLRDNGACVRSC 381
316 PLHNOEVTAEDGTQRCCKSKPCARVCYGLCMQYIKANSKFTIGTEL-----EPAGCK 369
362 PQDKMDKQGE-----CVPNGPCPKTC-----PGTVLAAGNIDSFNRCV 422
370 IFGSLAFPLESFDG--DPASNTA-----PLQEPQLVFTLEETITVYLISAMPDPLD 421
423 IDGNIRILDQTFSGQDVYAYNTMPRYIPLDPRRREVFSTVEITITVLYLIEGTHQFRN 482
422 LSVFONLOVIRGLIHNGAY-SLTLOGLISWGLRSLRELQSGALIIHNHLCFVHTV 480
483 LSYFNLLETIRGQLMESMFALAIVSSYSLEMRNLIKQISSGVVIOHNRDLCVASNI 542
481 PMDQCFRPHQALLHTARPEDEVGSLGCHQICAGHCGMGPGPTCCVACSGQFLGQEC 540
543 RWPALQKPEQKQWVWNELRADLCERKNTTICSDQCNEDGCGAGCTDCLTKRNFNNGTC 602
541 VEECRVLOGLPREYVNAHRLCPHRECOQPNQSVTCFGBPADQCVACAHYKDPFCVARC 600

```

```

Db      603 IADCGYISNAK--FDNFTCKI CHBECK-----TCNGAGADHCQCEVHVHDGQHCVSEC 654
Qy      601 P-----SGVK-----PDL 608
Db      655 PNKRYNDGVCRECHATCGCTGCPKDTIGIAGTTCNLAINDATYKRCCLLDKDPD- 713
Qy      609 SYMPIWKF--EDEBACQP-----CP-----NCH----- 632
Db      714 GY--FMEXVHPQEGSLKPLAGRAVCRKCHPLCCLTNYGYHEQVCSKCHYKRREOCET 721
Qy      633 -----SC-----VLDLDCG----- 641
Db      772 ECPADHYTDEBEGECFQRHPCNGCTGPGADCCSCNFKLFDANETGPRVYNTMPCNTS 831
Qy      642 -CPAEOR-----ASPLTS-----TSAVAGILLVAVGVFGI 673
Db      832 KCPLMRHNVNYOYTAIGPYCAASPRSSKITANLDVNMIFITITGAVLVPTICILCVTYI 891
Qy      674 LIQYIKANSKFIGIT-----ELPLPSGAMNQAMRILKETELRKXVVGSGAFGT 725
Db      892 CRQKQAKKEIVKXMTALSGCEDESEPLRPSNIGANLCRLVQDAELRKGGVGMGAFGR 951
Qy      726 YKGIWIPDGENVKIPVAIKVLRNTSPKANKELIDBAYVYAGVSPYRLLIGICTTST 785
Db      952 YKGVWVEGEGEVKIPVAIKELKSTGAESESEFLREAYIMASEHVNLLKLAVCMSSQ 1011
Qy      766 VQVLTQIMPYGCILDHVENRGRGLSGODLNMCMQIAKXSYLEDVNLVHRDLAARNVLY 845
Db      1012 MWLITQIMPICGLDLYVNNRDKIGSKALNWSQIAKXSYLEEKLVHRDLAARNVLY 1071
Qy      846 KSPNHVKITDGLARLDDIDETEHADGKVPFKWMALESILRRPFTHOSDWSYGVTVW 905
Db      1072 QTPSLVKITDGLAKLSSDSNEYKAAKGMPKMLALECIRNRFVFSKSDVAFFGTII 1131
Qy      906 ELMTFGAKPYDGIPIAREIPDLLENGERLPOPICTIVYIMWKMIDSECPRRRELY 965
Db      1132 ELTTFGRPHENIPAKDIPDLIEVGLKEQPEICSLDIYCTLLSCWMLDAMPTFKQLT 1191
Qy      966 SEFSMRADDPORFVYIQNEIDG--PASPBDSTFVRSLEDD--DMWDLDVADAEVLPQO 1020
Db      1192 TVFAEFARDPGRYLAIPGDKTRIPA-----YTSQDEKDLRKLAFTTDSGEAIAKPD 1244
Qy      1021 GFCEPDPAAGAGVNHHRSSSTRSGGDLTLGLEPSEEAR-----RSLPAPSEAG 1074
Db      1245 DYLOPKAPGPS-----HRTDCT-----DEMPKLNKCKDPGNKXSSG 1283
Qy      1075 SDVPDQ---DLQMGAAKGLQSLPTHDSPLQRYSEDPVLPBSTDGYVAPLTCSPQPEY 1131
Db      1284 DDERSSAREVGVGNLR-----LDLPVDEDDYIMP--TCQGPNN 1321
Qy      1132 VNCPDVRPQPSREGPLPAARPAATLERAKTLSPGKNGVVDVFAFGAVENPEYL-- 1189
Db      1332 NNNNN-----NPNQNNMAAVGVAAGY-----DLICVPSVUNPEYLLN 1360
Qy      1190 --TPGGAAPOPH-----PPAFSP-AFDNLVYWD 1216
Db      1361 AQLTGVGESPIFTQITIGIPWGGPTWEVVKVMPGSEPTSSDHEYND 1408

```

```

RESULT 12
ERBB ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (BC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses, Retroviridae; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]

```

```

RP      SEQUENCE FROM N.A.
RX      MEDLINE=85228222; PubMed=2988784;
RA      Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA      Crittenden L.B., Raines M.A., Kung H.-J.;
RT      "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT      processing and promoter insertion result in expression of an
RT      amino-truncated EGF receptor.";
RL      Cell 41:719-726 (1985).
CC      -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -1- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC      IN CHICKS WHICH HAVE BEEN INFECTED WITH THE AVIAN LEUKOSIS VIRUS
CC      AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC      THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC      C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC      -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC      PROTEIN.
CC      -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: M10066; AAA8763.1; ALT_INIT.
CC      PIR: A06643; TVCHLV.
CC      PIR: B06643; TVEVLV.
CC      HSSP: P1362; 1FEK.
CC      DR InterPro: IPR000719; Euk_Dkinase.
CC      DR InterPro: IPR001245; Tyr_Pkinase.
CC      Pfam: PF00069; Pkinase; 1.
CC      DR PRINTS: PR00109; TYRKINASE.
CC      DR ProDom: PD000001; Euk_kinase; 1.
CC      DR SMART: SMO0219; Tyrc; 1.
CC      DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
CC      DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
CC      DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
CC      KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC      KW Glycoprotein; Phosphorylation.
CC      FT DOMAIN 132 399
CC      FT NP BIND 138 146 ATP (BY SIMILARITY).
CC      FT BINDING 165 165
CC      FT ACT SITE 257 257 BY SIMILARITY.
CC      SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

```

```

Query Match 24.6%; Score 1666.5; DB 1; Length 634;
Best Local Similarity 50.7%; Pred. No. 1.9e-82;
Matches 358; Conservative 81; Mismatches 136; Indels 129; Gaps 19;

```

```

Qy      567 CAHYKDPFPCVAPCPGSKDLSYMPIWKEPDEBACQCPPICTHSCVLDLDCGPAPQ 646
Db      3 CAHFIDGHCYKAPAGVIGENDTL-WKXADANAVCQLCHPNCTRGCCKPGEGCP--- 58
Qy      647 RASPLTISVAVV-GILVVVLGVVFGIL--QYI---KANSKFIGITEL--PLTSGAM 698
Db      59 NSKTPSTIAGVVGGLCLVAVVIGIGLVRKRHYRKRTLRLLQRELVLPSTSGEA 118
Qy      699 PNOAMRILKETELRKXVVGSGAFGTVYGIWIPDGENVKIPVAIKVLRNTSPRANK 758
Db      119 PNOAHLRLKETELRKXVVGSGAFGTVYGIWIPDGENVKIPVAIKVLRNTSPRANK 178
Qy      759 ILDBAYVYAGVSPYRLLIGICTSTVOVLTQIMPYGCILDHVENRGRGLSGODLNM 818
Db      179 ILDBAYVYAGVSPYRLLIGICTSTVOVLTQIMPYGCILDHVENRGRGLSGODLNM 238
Qy      819 MOIAKXSYLEDVNLVHRDLAARNVLYKXSPNHVKITDGLARLDDIDETEHADGKVP 878
Db      239 VQIAKXSYLEDVNLVHRDLAARNVLYKXSPNHVKITDGLARLDDIDETEHADGKVP 298
Qy      879 KMALESILRRPFTHOSDWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPOPI 938

```

```

Db      299 KMALESLHRIYTHQSDVWSYGTVMELMTFGSKPYDGLIPASISSVLKGERLPOPP 358
Qy      939 CTIDVYIMVXCMWIDSECRPRFRELVSERFARMARDQRFVVIQ-NEDLGPASPLDSTFY 997
Db      359 CTIDVYIMVXCMWIDADSRPKFRELIAEFKMARDDPRYLVIQGERMMLPSPDTSKFY 418
Qy      998 RSLIEDDMGDLVAEELVPOQGFPCDPAPAGAGVNHHRHSSSTRSGGLTLGLER 1057
Db      419 RTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy      1058 SEEEAPRSPL-----APEGAGSDVFQDGLMGAKGLQSLPTHPSPLORYSEDPVPL 1112
Db      450 -----SRTPLLSLATSNNSATNCID-----RNGQGHPRVEDSFVGRYSSDPTGNR 496
Qy      1113 PSET--DGVAPLTCSPQPEYVNOQDVPRPQPPSRREGPLPAARPAAGTLERAKTLSPGKN 1170
Db      497 LEESIDDFL-----PAPEYVNO--LMPKPS-----TAMVON 527
Qy      1171 GVVKQVVF-----AFGAVENPEYLTPOGGAAPQHPPEPAPAFDNLVYW 1215
Db      528 QIYNNISLTAISKLEPMSDRYQNSHSTAYDNPEYL-----NTNQSPLAKTVFESSPYW 579

Qy      1216 DQ-----DPEF-----RGAPSTFKGTPTAENPEYIGLDVP 1246
Db      580 IQSGNHQINLNDMPDYQODPLNETKPNGLKVPAPENPEYLRAVAP 625

RESULT 13
ERBB_AVIER
ID ERBB_AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain E64).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
RX NCBI Taxid=79685;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RA MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
RT gene family.";
RL Cell 35:71-78(1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RA MEDLINE=84223957; PubMed=6328658;
RA Debut B., Henry C., Benaisa M., Bisette G., Clavierie J.-M.,
RA Saulie S., Martin P., Stehelin D.;
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
RT new type of oncogene.";
RL Science 224:1456-1459(1984).
CC -| CATLYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -| DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -| MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -| SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC Thin SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use, by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR      EMBL: K02006; AAA42394.1; ALT_INIT.
DR      EMBL: K01216; AAA42400.1; -.
DR      PIR: A00644; TVTH.
DR      HSSP: P11362; IEGK.
DR      InterPro: IPR000719; Euk_pkinase.
DR      InterPro: IPR001245; Tyr_pkinase.
DR      Pfam: PF00069; pkinase.1.
DR      ProDom: PD000001; Euk_pkinase.1.
DR      SMART: SM00219; TyrcKc.1.
DR      PROSITE: PS00107; PROTEIN_KINASE_ATP.1.
DR      PROSITE: PS00109; PROTEIN_KINASE_TYR.1.
DR      TRANSFAC: TS50011; PROTEIN_KINASE_DOM.1.
KW      Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW      Glycoprotein; Phosphorylation.
FT      DOMAIN 132 399
FT      NP_BIND 138 146 ATP (BY SIMILARITY).
FT      BINDING 165 165 ATP (BY SIMILARITY).
FT      ACT_SITE 257 257 BY SIMILARITY.
FT      CONFLICT 29 29 R -> W (IN REF. 2).
FT      CONFLICT 140 140 S -> F (IN REF. 2).
FT      CONFLICT 146 146 I -> V (IN REF. 2).
SQ      SEQUENCE 604 AA; 67633 MW; 76EBDD06745D609 CRC64;

Query Match 23.9%; Score 1620; DB 1; Length 604;
Best Local Similarity 50.5%; Pred. No. 5,7e-80;
Matches 348; Conservative 78; Mismatches 131; Indels 132; Gaps 18;

Qy      587 CAHYKDPFVCVAPRGVGVPLSYMPMKFPEDEGACQPCPINCSTHSCVLDJGCPAQ 646
Db      3 CAHFDGPHCVKACAPAGLGENDTL-VKRYADANAVCOLCHPCTRGCKPGLGCP 58
Qy      647 RASPLTSIVSAVV-GILLVVLGVVFGILI--QYI--KANSKFIGITEL--ELTPSGAM 698
Db      59 NGSKTPSIAGVAVGILLCLVYVGGIGILYRRRIYAKTLRLLCGRELVEPLTPSGEA 118
Qy      699 PNQAMRLKTELEARKYVJGSAFGVYVGMIPQSENVKIVAIKVRENSPANK 758
Db      119 PNQALRLKTEERKVKVJGSAFGYVGLMPBESEKVIPIAIKELKLEASPRANK 178
Qy      759 ILDEAYVAGVGSVYVSRILGICLTSTVQVLVQLMPYGCILLDHYRENGRLGSDLLNMC 818
Db      179 ILDENYVAVSDNVHVCRLGICLTSTVQVLITQMPYGCILLDVIREKNDIGSQYLLNMC 238
Qy      819 MOIAGKSYLEDVLRHDLAARVLYKSPNHVYKITDFGLARLLDIDETRYADGKVP 878
Db      239 VOIAGKMYLEERLVRHDLAARVLYKTPQHVYKITDFGLAKLLGADKEXYHAGGKVP 298
Qy      879 KMALESLHRRFTHQSDVWSYGTVMELMTFGKRPYDGLIPAREIPDLKGERLPOPP 928
Db      299 KMALESLHRIYTHQSDVWSYGTVMELMTFGSKPYDGLIPASEISSVLKGERLPOPP 358
Qy      939 CTIDVYIMVXCMWIDSECRPRFRELVSERFARMARDQRFVVIQ-NEDLGPASPLDSTFY 997
Db      359 CTIDVYIMVXCMWIDADSRPKFRELIAEFKMARDDPRYLVIQGERMMLPSPDTSKFY 418
Qy      998 RSLIEDDMGDLVAEELVPOQGFPCDPAPAGAGVNHHRHSSSTRSGGLTLGLER 1057
Db      419 RTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy      1058 SEEEAPRSPL-----APEGAGSDVFQDGLMGAKGLQSLPTHPSPLORYSEDPVPL 1112
Db      450 -----SRTPLLSLATSNNSATNCID-----RNGQGHPRVEDSFVGRYSSDPTGNR 496
Qy      1113 PSET--DGVAPLTCSPQPEYVNOQDVPRPQPPSRREGPLPAARPAAGTLERAKTLSPGKN 1170
Db      497 LEESIDDFL-----PAPEYVNO--LMPKPSYAM----- 524
Qy      1171 GVVKQVVF-----GAVENPEYLTPOGGAAPQHPPEPAPAFDNL 1211
Db      525 -VQNDYVFIPLTALSKLEPMSDRYQNSHSTAYDNPEYL-----NTNQSPLAKTVFES 575
Qy      1212 LYWDQDPPERGAPPSFKGTPTAENPEY 1240

```

```

Db      576 SPYWIGSGNHQ-----INLNDPY 594
RESULT 14
ERBB_AVIEU STANDARD; PRT; 540 AA.
AC P11273.
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ts167).
OC Viruses; Retroid. viruses; Retroviridae; Avian type C retroviruses.
OK NCBI_TaxID=103898;
RN SEQUENCE FROM N.A.
RP MEDLINE=87064458; PubMed=2878364;
RA Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RT "A single amino acid substitution in v-erbB confers a thermolabile
RT phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT cells."
RL Mol. Cell. Biol. 6:1751-1759 (1986).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M13179; AAA42401.1; -.
DR PIR; A25231; TVPEVB.
DR HSSP; P11362; IEGK.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00069; Pkinase; 1.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
DR Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KM Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT VARIANT 270 270 H -> D (IN THE PROLAMILE V-ERBB).
SQ SEQUENCE 540 AA; 60412 MW; 5B53297AA06B65D CRC64;
Query Match 22.8%; Score 1547; DB 1; Length 540;
Best Local Similarity 53.2%; Pred. NO. 4.2e-76;
Matches 329; Conservative 71; Mismatches 122; Indels 96; Gaps 16;

```

```

Db      179 ILDEAYVNASVDNHNVCRLIGLICITSTVOLITOLMPYGCCLDIYIRHKONIGSQYLLNMC 238
MoIAKMSYEDVLRHDLAARVNLVKSNNHYKTFDGLARLLDIDETRYHADGKVP 878
VQIAKGMVLEERHMYRDLAARVNLVKTQHKITDFGLAKQLGADKRYHAGKVP 238
KMALESLIRRRFTHQSDWMSYGVTELMTFGAKPYDGIIPAREIPDLLEKGRLLPQPI 938
KMALESLIRHRYTHQSDWMSYGVTELMTFGSKPYDGIIPASEISSVLEKGRLLPQPI 358
CTIDVYMMVYCNWIDECRPFELVSEFSRNARDQREVVO-NEDLGAPASPLDSTFY 997
CTIDVYMMVYCNWIDECRPFELVSEFSRNARDQREVVO-NEDLGAPASPLDSTFY 997
RSLLEDDMDLVDAEEYLVPOQGFPCPDAPGAGVHHRRSSSTRSGGDLTLGLEP 1057
RTLMEEDMEDIVADAEYLVHQGF-----NSPST----- 449
1058 SEEARPSPL-----ASEGAGSDVFPGDLGMGAAGLQSLPHDPSPLORYSEDPVPL 1112
450 -----SRTPLLSSLSATSNNSATCIDRNQ-----H----- 476
QY PSETDGYVAPLTCSPQEEYVNPQPVPRQPPSPREGPLAARPAAGT-LERAKTLPCKG 1171
Db      477 PVREDGL-----PAREYVQ--LWKKESTAVQNYVISTALISKLPIDRYQ-- 527
QY 1172 VKQVFAFGAVENPEYL 1189
Db      528 -----SHSTAVDPEYL 539
RESULT 15
EGFR_CHICK STANDARD; PRT; 703 AA.
AC P13387;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER
DE (fragment).
GN EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauromia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OX NCBI_TaxID=9031;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=88261272; PubMed=3260329;
RA Lax I., Johnson A., Hawk R., Sap J., Bellot F., Winkler M.,
RA Ullrich A., Vennstrom B., Schlessinger J., Glavol D.,
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha."
RT Mol. Cell. Biol. 8:1970-1978 (1988).
RL Mol. Cell. Biol. 8:1970-1978 (1988).
CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF.
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

or send an email to license@isb-sib.ch.

```
CC -----
DR EMBL; M20386; AAA48760.1;
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Dkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_Dkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SMO0261; FU; 4.
DR PROSITE; PS00107; PROTEIN KINASE_ATP_PARTIAL.
DR PROSITE; PS00109; PROTEIN KINASE_TYR_PARTIAL.
DR PROSITE; PS00011; PROTEIN KINASE_DOM_PARTIAL.
KM Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KV Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 30
FT CHAIN 31 >703
FT DOMAIN 31 654
FT TRANSMEM 655 667
FT DOMAIN 668 >703
FT DISULFID 197 206
FT DISULFID 201 214
FT DISULFID 222 230
FT DISULFID 226 238
FT DISULFID 239 247
FT DISULFID 243 255
FT DISULFID 258 267
FT DISULFID 271 298
FT DISULFID 302 314
FT DISULFID 318 333
FT DISULFID 336 340
FT DISULFID 513 522
FT DISULFID 517 530
FT DISULFID 533 542
FT DISULFID 546 562
FT DISULFID 565 581
FT DISULFID 569 589
FT DISULFID 592 601
FT DISULFID 605 627
FT DISULFID 630 638
FT DISULFID 634 646
FT CARBOHYD 134 134
FT CARBOHYD 190 190
FT CARBOHYD 200 200
FT CARBOHYD 200 200
FT CARBOHYD 359 359
FT CARBOHYD 368 368
FT CARBOHYD 420 420
FT CARBOHYD 573 573
FT CARBOHYD 578 578
FT CARBOHYD 613 613
FT CARBOHYD 633 633
FT CARBOHYD 648 648
FT NON_TER 703
SQ SEQUENCE 703 AA; 77427 MW; AFE2DE11B735A690 CRC64;

Query Match      22.3%; Score 1512; DB 1; Length 703;
Best Local Similarity 43.0%; Pred. No. 4,2e-74;
Matches 304; Conservative 113; Mismatches 254; Indels 36; Gaps 16;
```

```
QY 8 RWGLLALIPGAA-----STQVCTGIDMKLRLPASPEHLDMLRHLKYQQCVVQGNLE 61
DB 13 RGAAYVLLLLGLVALCSAVEEKVCQGINNKLTLQGHVEDHFTSLQRMYNCEVVLNLE 72
QY 62 LTYLPTNASLFLDIOEVGVVLAHNOVRQVPLQRLRIYRGQLPEDNYALAVLDNGD 121
DB 73 IYVEHNRDLTLFKTIQEVAGVLLALNMVDVIFLENIQIRGNVLDNSFALAVLSNYH 132
QY 122 PLNNTPVTGASPGSLRELQLSLTEILKGGVLIQRPOLCYODTIWKDI FHKRNQAL 181
DB 133 -MNKTQ-----GLRELPMKRLSEIILNGVKISNNPKLCKMPTVLMDIIDTSRK-PL 182
QY 182 TLID-TNRSRACHPSPMCKGSRCKGSESEDCQSLLTRTVCAAGCA-RCKGPLPTDCQHEQ 239
```

```
DB 183 TLVDFASNLSSCKPCHENCTEDHGMWGAEGNQCLTKTVI CAQQCSGRGRGVYSDCCHNQ 242
QY 240 CAACCTGPKRSDCLACLFHNSGICELHCAIYNTDPEFSPNPEGRYTGASCVTAC 299
DB 243 CAACCTGPRASDCLACKFRDDATCKQTCPLVLPYPTTYQMDVNEGKYSFGATVREG 302
QY 300 PNYLSTDVSGCTLVCPILHNOEYTAEDGTQRCCKSKPCARVYGVGMQYIKANSKFIGI 359
DB 303 PHNVVTFDHGSCVWASGCTDTYEV-EEGVYKCKCKCGQLCSKVCNGIGIGELKILS-INA 360
QY 360 TELE-FAGCKKIFGSLAFLPESFDGDPASNTALQEQLOVEFTEILEITGYLYISWPOS 418
DB 361 TNDSFKKCKTKINDVSLPVAFLGAFKTLPLDPKLDVFRTVIEISGFLLIQAMPDN 420
QY 419 LPDVSFQNLQVIRGILHNGAVSLTQGLGISMGLSRLSELGSLALIHNTHLCFVH 478
DB 421 ATDLAENLEIIGRTKQGGYSLAVNVLKISGLRSIKE-SDDDIAIMKNKNCYAD 480
QY 479 TVPDQLFRNPHQALLHTANRPEDECVGBGLACHQLCARGHCKGPGFTQCVCNSQFLRQ 538
DB 481 TMNRSFLFATQSQTKIIOQRNKNCDTADRHVCDPLCSVDGCMGPGPFHCFSCRFPSRK 540
QY 539 ECVEECVLQGLPREYVNAHCLPCHPECCPONG---SVTCFGEADCCVACAHYCDPF 595
DB 541 ECVRQCNILQGEPRFERDSKCLPCHSECLVQNSTAVNTTCSGPGCDHCKKCHFIDGR 600
QY 596 CVARCPGSKPEKDLSPYPIWKFPPEBGAQCPPINCHSCVLDLCKGCPAEQRPASPLTSIV 655
DB 601 CVKACPGVJGENDTL-VMKYADANAVCQCHPNCRGCKGPGLEGCP--NNSKTFPSIA 656
QY 656 SAVY-GILLVVVLGVVFGILI--QYI--KANSKFIGITL--PLRP 694
DB 657 AGVVGILCLVVGILGILYLRHHIVRKRRLRLLOERLVEPLRP 703
```

Search completed: July 22, 2003, 09:18:11
Job time : 16.5741 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:01:42 ; Search time 46.5261 Seconds
(without alignments)
5522.503 Million cell updates/sec

Title: SEQ4-653-675-12
Perfect score: 6776
Sequence: 1 MELALACRMGLIALLPQA.....TEKGTPTAENPEYLGLDVPV 1247

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPTRMBL_21:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6095	89.9	1259	6 018735	018735 canis fami
2	3057	44.1	1209	11 090X70	090X70 rattus norv
3	3030	44.7	1210	11 09EP98	09EP98 mus musculu
4	2646	39.0	1165	13 09YH40	09YH40 xiphophorus
5	2643.5	39.0	1137	13 09W656	09W656 gallus gall
6	2262	33.4	1328	13 P79754	P79754 fugu rubrip
7	2007.5	29.6	1433	5 08BIH9	08BIH9 anopheles g
8	1871	27.6	419	4 09UK79	09UK79 homo sapien
9	1739	25.7	367	11 08R2X1	08R2X1 mus musculu
10	1697.5	25.1	412	4 08WYV0	08WYV0 homo sapien
11	1637	24.2	729	15 086712	086712 avian trou-
12	1635	24.1	567	15 086714	086714 avian trou-
13	1570.5	23.2	962	15 064895	064895 avian eryth
14	1562	23.1	545	15 085468	085468 avian eryth
15	1506.5	22.2	655	11 09WV55	09WV55 mus musculu
16	1490.5	22.0	643	11 09ERV6	09ERV6 mus musculu

17	1262	18.6	1193	5 09Y1X8	09Y1X8 ephydria f
18	1183.5	17.5	1368	5 023821	023821 caenornabdi
19	1144	16.9	1717	5 026566	026566 schistosoma
20	1125	16.6	527	13 090836	090836 gallus gall
21	1001.5	14.8	478	11 09ES80	09ES80 rattus norv
22	942.5	13.9	539	13 09PSH2	09PSH2 gallus gall
23	906	13.4	165	4 014256	014256 homo sapien
24	887	13.1	176	11 0923V5	0923V5 rattus norv
25	806.5	11.9	346	13 P11776	P11776 xiphophorus
26	778	11.5	435	5 08SZM1	08SZM1 drosophila
27	754.5	11.1	311	13 091562	091562 xiphophorus
28	743.5	11.0	1352	13 09PVZ4	09PVZ4 xenopus lae
29	734	10.8	331	4 09BUD7	09BUD7 homo sapien
30	727	10.7	1671	5 09NIV5	09NIV5 biophalaria
31	723	10.7	149	6 09B666	09B666 oryctolagus
32	698	10.3	1418	13 093457	093457 scophthalmu
33	688.5	10.2	1368	13 08UM85	08UM85 paralicthy
34	676.5	10.0	1369	13 08UM86	08UM86 paralicthy
35	672	9.9	1412	13 08UM84	08UM84 paralicthy
36	668	9.9	1358	13 073798	073798 xenopus lae
37	665.5	9.8	1472	5 09USAB	09USAB bombyx mori
38	647.5	9.6	1418	13 08UM83	08UM83 paralicthy
39	642	9.5	1245	13 09YGH8	09YGH8 scophthalmu
40	634	9.4	1371	11 09QVW4	09QVW4 rattus sp.
41	615.5	9.1	2144	5 09VD94	09VD94 drosophila
42	595	8.8	987	11 091YMO	091YMO mus musculu
43	591.5	8.7	1091	4 09UMQ4	09UMQ4 homo sapien
44	589	8.7	1036	4 007912	007912 homo sapien
45	584	8.6	987	11 099MR2	099MR2 mus musculu

ALIGNMENTS

RESULT 1	
018735	PRELIMINARY; PRT; 1259 AA.
AC 018735	
DT 01-JAN-1998	(TREMBLrel_05, Created)
DT 01-JAN-1998	(TREMBLrel_05, Last sequence update)
DT 01-JUN-2002	(TREMBLrel_21, Last annotation update)
DE	Erdb-2.
OS	Canis familiaris (Dog).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX	NCBI_TaxID=9615;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Yokota H.;
RT	"CDNA cloning of erdb-2 from canine mammary gland.";
RL	Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AB008451; BAA23127.1; -
DR	HSSP; P11362; ITCG.
DR	InterPro; IPR002048; EF-hand.
DR	InterPro; IPR000494; EGFR_L_domain.
DR	InterPro; IPR000719; Euk_Pkinase.
DR	InterPro; IPR002174; Purin-like.
DR	InterPro; IPR001245; Tyr_Pkinase.
DR	InterPro; IPR004019; YLP_motif.
DR	Pfam; PF00757; Pkinase_1.
DR	Pfam; PF00069; Pkinase_1.
DR	Pfam; PF01030; Recep_L_domain; 2.
DR	Pfam; PF02757; YLP; 2.
DR	ProDom; PD000001; Euk_Pkinase; 1.
DR	SMART; SMO0261; FU; 3.
DR	SMART; SMO0219; Tyrc; 1.
DR	PROSITE; PS00107; EF_HAND; UNKNOWN 1.
DR	PROSITE; PS00011; PROTEIN_KINASE_ATP; 1.
DR	PROSITE; PS00109; PROTEIN_KINASE_DOM; 1.
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW	ATP-binding; Transferase; Tyrosine-protein kinase.
SQ	SEQUENCE 1259 AA; 137989 MW; E37364D49C4CD46 CRC64;

Query Match		89.8%; Score 6095; DB 6; Length 1259;
Best Local Similarity		89.5%; Pred. No. 0;
Matches 1128; Conservative		46; Mismatches 72; Indels 14; Gaps 5;
Qy	1	MELALCRWGLLLALPPGAASSTOVCTGDMKRLPASPEETHIDMLRHLYOGQVVGNTL 60
Db	1	MELAMCRWGLLLALLPSGAAGTGVCTGIDMKRLPASPEETHIDMLRHLYOGQVVGNTL 60
Qy	61	ELTYLPTNASLSFQDIQEVQGYLLAHNOVROVPLORLRYVGTQGFEDNALAVLDNG 120
Db	61	ELTYLPANASLSFQDIQEVQGYLLAHNSQVROIPQRLRYVGTQGFEDNALAVLDNG 120
Qy	121	DPLNNTPTVGTASFGURELQBSLEILKGYLIQRPQLCYQDTILMKDIFHNQOLA 180
Db	121	DPLEGGIPAPGAAGGURELQBSLEILKGYLIQRPQLCYQDTILMKDIFHNQOLA 180
Qy	181	LTLDITNRSRACHCSPCKGSRCKWGSSEDCOSLRITVAGGACRCKPLPTDCHEOC 240
Db	181	LTLDITNRSACHCSPCKGSRCKWGSSEDCOSLRITVAGGACRCKGPPTDCHEOC 240
Qy	241	AAGCTGPRGSDCLALPHNHSIGICELCPALVTYNTDTESMNPBGRYTFGASCTYACP 300
Db	241	AAGCTGPRGSDCLALPHNHSIGICELCPALVTYNTDTESMNPBGRYTFGASCTYACP 300
Qy	301	YNYLSTDVSGCTLVCPILHNOVTAEDTQCEKSKRCARVCTGLGMQYIKANSKPIGIT 360
Db	301	YNYLSTDVSGCTLVCPILHNOVTAEDTQCEKSKRCARVCTGLGMQYIKANSKPIGIT 360
Qy	361	ELFAPGCKIFGSLAFPESEFDGPASNTAPLOPEQLQVETLEETGYLYISAMPDSL 420
Db	361	IQEFAGCKIFGSLAFPESEFDGPASNTAPLOPEQLQVETLEETGYLYISAMPDSL 420
Qy	421	DLVSFQNLQVIRGILHNGAYSLTLQGLISWTGLNSRLPSLGGALLIHNTHLCVHTV 480
Db	421	NLSVFQNLQVIRGVLHDGAYSLTLQGLISWGLNSRLPSLGGALLIHNTHLCVHTV 480
Qy	481	PMDQLFNRPHQALHTANRPEDECVEGLACHQICARGHGMGSPQVCNCSQFLRGQCC 540
Db	481	PMDQLFNRPHQALHTANRPEDECVEGLACHQICARGHGMGSPQVCNCSQFLRGQCC 540
Qy	541	VEECRVLOGLPREYVNAHCLPCHPECCOPNGSVTCGPADOCVACAHYKDBPFCVAC 600
Db	540	VEECRVLOGLPREYVNAHCLPCHPECCOPNGSVTCGPADOCVACAHYKDBPFCVAC 600
Qy	601	PSGVKPDLSYMIKPFDEBEGACQPCINCTHSCVLDLDCGCAEGRASPISIVAAVG 660
Db	600	PSGVKPDLSYMIKPFDEBEGACQPCINCTHSCADDEGCAEGRASPISIVAAVG 659
Qy	661	ILLVVLGVVFGILI---QYIKANS--KEIGITEL--PLTPSGAMPNOAQRILKETEL 712
Db	660	ILLAVVGLVGLIIRKRRQKIRKYTWRRLLQETELVEPFTPSGAMPNOAQRILKETEL 719
Qy	713	RKVAVLGGAGTGYKKIWIIPDGNVYKIPAIKYLRENTSPKANKLILDEAYMAVAGVSP 772
Db	720	RKVAVLGGAGTGYKKIWIIPDGNVYKIPAIKYLRENTSPKANKLILDEAYMAVAGVSP 779
Qy	773	YVSRLLGICLTSTVQLTQMLPQYCLLDHVENRGRGLSGODLNMCMQIAKGSYLEDR 832
Db	780	YVSRLLGICLTSTVQLTQMLPQYCLLDHVENRGRGLSGODLNMCMQIAKGSYLEDR 839
Qy	833	LVHRDLAARNVLVSPNHVKITDGLARLLDIDETEHADGCVPIKMMALLESILRRRT 892
Db	840	LVHRDLAARNVLVSPNHVKITDGLARLLDIDETEHADGCVPIKMMALLESILRRRT 899
Qy	893	HOSDVMASGYVMEIMTFGAKPYDGIAREIPDLLEKGEELPPPICTIDVYIMKCM 952
Db	900	HOSDVMASGYVMEIMTFGAKPYDGIAREIPDLLEKGEELPPPICTIDVYIMKCM 959
Qy	953	IDSECRPREFELVESFRMADPQRFVVIQNEDELGPASPLDSTFYSSLLEDDMGDLVDA 1012
Db	960	IDSECRPREFELVESFRMADPQRFVVIQNEDELGPASPLDSTFYSSLLEDDMGDLVDA 1019
Qy	1013	EFLVLPQGGFFCPRDAPGAGMVHHRSSSTRSGGDLTLGLEPBEERAPSLAPSG 1072

Db	1020	EFLVLPQGGFFCPRDAPGAGTARHRRSSSTRNGGSLTLGLEPSEEPSPKSLAPSG 1079
Qy	1073	AGSDVPBGDLGMAAAGLQSLPHDSSPIQRFSEPTTVLPSSTGYVAPLTCSPQPEV 1132
Db	1080	AGSDVPBGDLGMAAAGLQSLPHDSSPIQRFSEPTTVLPSSTGYVAPLTCSPQPEV 1139
Qy	1133	NQPVPRQPSPEGLPAPAPGATLER----AKTLSPKXGVKGVDFARGAVENTE 1187
Db	1140	NQPVPRQPSPEGLPAPAPGATLER----AKTLSPKXGVKGVDFARGAVENTE 1199
Qy	1188	YLTPQGAAPQHPHPAPAFSPADNLVYWDQDPERGAPSTGKTPTAENPEYLGIDPV 1247
Db	1200	YLAPRGAAPQHPHPAPAFSPADNLVYWDQDPERGAPSTGKTPTAENPEYLGIDPV 1259
RESULT 2		
ID	Q9GX70	PRELIMINARY; PRT; 1209 AA.
AC	Q9GX70	
DT	01-MAY-2000 (TREMBlrel. 13, Created)	
DT	01-MAY-2000 (TREMBlrel. 13, Last sequence update)	
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)	
DE	Epidermal growth factor receptor.	
GN	EGFR.	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OK	NCBI_Taxid=10116;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
EX	MEDLINE=9025888; PubMed=2342466;	
SA	Barp H.S.; Harris J.; Raymond V.W.; Blaasband A.J.; Lee D.C.;	
SA	"A truncated, secreted form of the epidermal growth factor receptor is	
RT	encoded by an alternatively spliced transcript in normal rat tissue."	
RT	Mol. Cell. Biol. 10:2973-2982(1990).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
SA	Barp H.S.;	
SA	Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
SA	Guttridge K.; Dawson T.L.; Barp H.S.;	
SA	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; M37394; AAF14008.1; -	
DR	HSSP; P13362; IRGK.	
DR	InterPro; IPR000494; EGFR_L domain.	
DR	InterPro; IPR000719; Euk_Pkinase.	
DR	InterPro; IPR002174; Furin-like.	
DR	InterPro; IPR001245; Tyr_Pkinase.	
DR	Pfam; PF00757; Furin-like; 1.	
DR	Pfam; PF00069; pkinase; 1.	
DR	Pfam; PF01030; Recep_L domain; 2.	
DR	PRINTS; PR00109; TYRKINASE.	
DR	ProDom; PD000001; Euk_Pkinase; 1.	
DR	SMART; SM00261; FU; 3.	
DR	SMART; SM00219; TYRK; 1.	
DR	PROSITE; PS00107; PROTEIN KINASE ATP; 1.	
DR	PROSITE; PS0011; PROTEIN KINASE_DOM; 1.	
DR	PROSITE; PS00109; PROTEIN KINASE_TYR; 1.	
KW	ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.	
SO	SEQUENCE 1209 AA; 134891 MW; 96FEETFECCIB773 CRC64;	
Query Match		45.1%; Score 3057; DB 11; Length 1209;
Best Local Similarity		49.2%; Pred. No. 1.2e-220;
Matches 625; Conservative		171; Mismatches 359; Indels 115; Gaps 29;
Qy	11	LLALLPQGA--STQVCTGDMKRLPASPEETHIDMLRHLYOGQVVGNTLPTN 68